DOCUMENT RESUME

ED 337 144 IR 015 194

AUTHOR Winkler, John D.; Polich, J. Michael
TITLE Effectiveness of Interactive Videodisc in Army Communications Training.
INSTITUTION Rand Corp., Santa Monica, Calif.
SPONS AGENCY Office of the Assistant Secretary of Defense for Force Management and Personnel (DOD), Washington, DC.
REPORT NO ISBN-0-8330-1057-3; RAND/R-3848-FMP
PUB DATE Nov 90
CONTRACT MDA903-85-C-0030
NOTE 90p.
PUB TYPE Information Analyses (070) -- Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Communications; Comparative Analysis; *Computer Assisted Instruction; Computer Simulation; Conventional Instruction; *Evaluation Methods; Instructional Effectiveness; *Interactive Video; *Intermode Differences; Microcomputers; Military Training; Optical Data Disks; Postsecondary Education; Technical Education; *Training Methods; Transfer of Training

ABSTRACT

This report presents the results of RAND research conducted at the U.S. Army Signal Center, Fort Gordon, Georgia, to evaluate the effectiveness of an interactive videodisc (IVD) system used to facilitate training in a variety of military occupational specialities. The objectives of the study were to: (1) develop a methodology for assessing the instructional effectiveness of IVD technology, which links a microcomputer and laser videodisc to provide interactive instruction with high-resolution video displays; (2) apply the methodology to evaluate the benefits of an IVD training system used in communications training; and (3) provide a general model for assessing related training technologies in a broad range of courses and environments throughout the defense community. This report documents two RAND studies which compared the effects of hands-on equipment training with IVD training to determine the effectiveness of IVD in specific types of training and the transfer of that training to new situations. Discussions of issues raised by these studies in regard to how IVD technology may be used most appropriately in other training applications and training conditions conclude the report. (56 references) (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 362 144

Effectiveness of Interactive Videodisc in Army Communications Training

John D. Winkler, J. Michael Polich

IR 015194

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Janet Valadez

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

RAND | NATIONAL DEFENSE RESEARCH INSTITUTE

The research described in this report was sponsored by the Assistant Secretary of Defense (Force Management and Personnel). The research was conducted in the National Defense Research Institute, RAND's federally funded research and development center supported by the Office of the Secretary of Defense, Contract No. MDA903-85-C-0030.

ISBN: 0-8330-1057-3

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of The RAND Corporation do not necessarily reflect the opinions or policies of the sponsors of RAND research.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

R-3848-FMP

Effectiveness of Interactive Videodisc in Army Communications Training

John D. Winkler, J. Michael Polich

November 1990

Prepared for the
Assistant Secretary of Defense
(Force Management and Personnel)

RAND

Approved for public release; distribution unlimited

PREFACE

This report presents the results of RAND research conducted at the U.S. Army Signal Center, Fort Gordon, Georgia, to evaluate the effectiveness of an interactive videodisc (IVD) system used to facilitate training in a variety of military occupational specialties. The objectives of the study are to develop a methodology for assessing the training effectiveness of IVD technology, apply the methodology to evaluate the benefits of an IVD training system used in communications training, and provide a general model for assessing related training technologies in a broad range of courses and environments throughout the defense community. The results should be of interest to manpower and training analysts and policymakers in the Office of the Secretary of Defense and in the Army, as well as to personnel in the military services who are contemplating developing interactive videodisc hardware and courseware for training purposes.

This research was performed in the Defense Manpower Research Center, part of RAND's National Defense Research Institute, an OSD-sponsored federally funded research and development center. The research was sponsored by the Assistant Secretary of Defense (Force Management and Personnel), in cooperation with the U.S. Army Signal Center and the Defense Training and Performance Data Center (TPDC).

An earlier version of the summary of this report appeared in a Department of Defense report to Congress, "The Potential of Interactive Videodisc Technology for Defense Training and Education," 1989.

SUMMARY

As the technical sophistication of military weapon and support systems has increased, the services have sought new ways to use technology to train for more complex tasks. Prominent among new training technologies is interactive videodisc (IVD) technology, which links a microcomputer and laser videodisc to provide interactive instruction with high-resolution video displays. This report documents two RAND studies of Army IVD applications, employing rigorous experimental designs and post-experimental performance assessments to evaluate the effects of alternative uses of IVD in Army communications training.

BACKGROUND

Defense modernization has brought complex new weapon and support systems into the inventories of the military services. Improvements in military technology, however, bring conflicting pressures on the services' training establishments. The growing variety and complexity of many new systems tend to raise skill requirements, leading to pressures for longer and more expensive training courses. At the same time, some operational equipment has become so costly that the training base can at best afford only a few pieces that resemble those actually used in the field. In field units themselves, where equipment is available, it is difficult to ensure standardization and quality of training.

Military trainers have begun to respond to such challenges by expanding their use of new computer-based, visually oriented training devices and simulators. These technologies have the potential to simulate a variety of new equipment, provide individualized yet standardized instruction, engage learners in dynamic problem-solving situations, and provide immediate feedback about performance. Among recent innovations, interactive videodisc technology, which consists of an integrated microcomputer, video display, laser videodisc, and instructional software (termed interactive courseware), represents a new training device with considerable promise.

The U.S. Army Signal Center at Fort Gordon, Georgia, has pioneered the use of IVD systems for training soldiers in a variety of communications-electronics military occupational specialties (MOSs). Signal Center developers of an early IVD system—a predecessor to

6

the Army's Electronic Information Delivery System (EIDS), for which a substantial acquisition began in Fiscal Year 1988—have hypothesized that school-based IVD training may increase student proficiency and reduce hands-on training requirements for a broad range of specialties. They also believe that IVD systems have potential in field units for refresher and on-the-job training. Demonstration of these hypothesized benefits could affect decisions by the Army and by the other services about purchasing EIDS hardware, developing interactive courseware, and allocating EIDS training systems across various specialties and environments.

The possible benefits of IVD technology are of interest not only to the services, but also to the Office of the Secretary of Defense (OSD), which has oversight responsibility for the efficiency of training. OSD interest in the Signal Center experience, and in the effectiveness of IVD more broadly, has been heightened as the various services have become interested in applications of interactive videodisc and similar technologies. However, to date the systematic data needed to assess the potential benefits of IVD training have been lacking. To provide such data and to establish a model for future research in this area, RAND undertook a series of studies of IVD in cooperation with the Signal Center and OSD's Defense Training and Performance Data Center. This report presents the results of these studies and their implications for uses of IVD in military settings.

OBJECTIVES AND APPROACH

The objectives of this study were to develop a methodology for assessing the benefits of innovative training technologies, to apply the methodology to evaluate the effectiveness of an IVD training system used at the Army Signal Center, and to define general conditions for effective use of IVD technology.

Our approach applied principles of controlled experimentation to compare effects of alternative methods used to train equivalent groups of soldiers. We report the results of two studies. In both, the effects of traditional hands-on equipment training (the control condition) are compared with effects of a training regimen using IVD (the experimental condition). The experimental and control groups were formed using a statistical randomization model developed at RAND that provides a close match between groups on such factors as aptitude, educational background, demographic characteristics, and military experience. The training received by each group was carefully monitored, and the effects of alternative training methods were com-

pared using multivariate analysis of objective, training- and job-related performance criteria.

The two studies examine two common applications of IVD in the Army: as a device used to *supplement* or augment existing hands-on training, and as a device used to *simulate* or replace hands-on equipment training. The first use of IVD increases training opportunity while increasing costs; the second use of IVD can maintain existing training opportunity while decreasing costs. The studies provide empirical evidence of IVD effectiveness in the specific MOSs trained, and they point to implications for IVD training policy in many other military settings.

SUPPLEMENTARY TRAINING WITH IVD: MOS 31M

The initial experiment evaluated the effects of IVD on student proficiency when used as a device to supplement hands-on equipment training in MOS 31M, Multichannel Communications Equipment Operator. The experimental training took place during two weeks of the course when students learned to install "low-capacity" radio equipment (AN/TRC-145). The experiment lasted seven months and covered 428 active duty trainees who were assigned to one of two groups. The control group received hands-on training at installation using only radio assemblages, whereas the experimental group received both hands-on training with radio assemblages and IVD training. Each group had an equal number of radio assemblages available (normally 10 assemblages for a class of up to 25 students). In the experimental group, the IVD provided an additional eight training positions to allow trainees more opportunity to practice radio-related tasks within the allotted time.

Several weeks later, the performance of each trainee at assemblage installation was assessed using the Reactive Electronic Equipment Simulator (REES), a high-fidelity, computer-controlled facility that contained the pertinent radio assemblages. The REES computer provided data on the accuracy with which trainees accomplished the installation, as well as the amount of time and effort required to successfully install the radio assemblage. Trainees' job knowledge was also assessed using a written examination, which contained elements of job knowledge that were trained as well as measures of trainees' attitudes toward the training that they received.

The research hypothesis in this study was that IVD use would increase the efficiency of training while improving student proficiency. Results showed that the IVD was extensively implemented in the ex-

perimental classrooms; the addition of IVD to the classroom led to a 45 percent increase in time spent practicing installation of radio assemblages. Thus, those students received increased training opportunity without lengthening their overall amount of time in the course. In this respect, the use of IVD allowed instructors to make more efficient use of student time.

IVD training also increased soldier proficiency, as assessed in the high-fidelity simulator. Regression analyses showed that supplemental IVD training caused statistically significant reductions in the time needed to install the radio equipment, the number of trials (amount of effort) needed to accomplish the installation, and the likelihood of a student error during the installation process. These reductions were modest, however, ranging between 10 and 20 percent.

SUBSTITUTION TRAINING WITH IVD: MOS 31Q

The second experiment examined the effects of substituting IVD technology for more expensive equipment in MOS 31Q, Tactical Satellite/Microwave Systems Operator. This experiment, lasting 10 months and encompassing 336 trainees, focused on training the alignment and adjustment of complex and expensive tropospheric scatter (TROPO) radio assemblages. The approach held the amount of training opportunity constant while varying the resources used for training. Students were assigned to one of two groups: Half carried out exercises in a classroom equipped with seven TROPO radios and eight closely related line-of-sight (LOS) radios, while the other half carried out similar exercises in a classroom that contained only one of each type of radio but had eight IVD units—a much less expensive complement of training devices.

Immediately after the training, we assessed the performance of each trainee using a hands-on test based on the Army *Soldiers Manual*, including three relevant tasks [intermediate frequency (IF) gain alignment, automatic gain control (AGC) alignment, and squelch adjustment]. The hands-on tests were administered by objective assessors, trained and monitored by RAND, who were unaware of how each soldier had been trained. For each test, we determined whether the trainee could accomplish each of the tasks within the respective Army time standard, and we recorded errors made during task performance. Trainees also received a written test providing measures of task knowledge and attitudes toward the training that they received.

For this study, the research hypothesis was that students would be equally proficient at the tasks, whether they were trained under the traditional equipment-only regimen or under the alternative regimen in which IVD was used at a substantial saving in training resources. Our analyses confirm the hypothesis for measures of proficiency and job knowledge. As an illustration, we summarize results for performance on the IF gain alignment, the most difficult of the tasks. The results show that students used IVD extensively in the experimental classroom, accomplishing 58 percent of their training sessions on IVD. Students in the control group received approximately the same number of training sessions, but of course 100 percent of their training was done on actual equipment. Despite this substitution, the performance of the groups on the hands-on test was statistically indistinguishable.

Our analyses show similar results for student performance on AGC alignment and squelch adjustment—ability to accomplish the task was the same, whether students were trained with actual equipment or with a mix of IVD and actual equipment. However, for these tasks the IVD-trained students appeared slightly more likely to make procedural errors, and they were less satisfied with the training they received.

CONCLUSIONS

The results of these experimental studies show that IVD technology can be beneficial in its two most common types of application: as a supplement to existing training or as a substitute for more expensive training resources. In MOS 31M, the addition of IVD provided increased training opportunity and caused improvements in measures of subsequent task proficiency.[1] In MOS 31Q, the replacement of some equipment training with IVD training did not diminish students' ability to perform the relevant tasks. The studies thus confirm many of the benefits of IVD technology espoused by its advocates, at least for the applications we examined.

At the same time, however, information collected in both studies suggested some important conditions that may affect when and where one chooses to use IVD. In MOS 31M, our data showed that most trainees received ample hands-on training opportunity, even in the control group where instructors perceived an equipment shortage; in

[1]Other research studies suggest that the additional practice provided by such training technologies can permit a reduction in the allotted training time.

fact, nearly all trainees were eventually able to perform the installation successfully. We speculated (without strong statistical evidence) that frequent practice on real equipment had given most students a fairly high level of basic proficiency, which may have limited the benefits that could be gained by adding IVD. If correct, this suggests an important criterion for using IVD as a supplement to existing training resources: Supplementation is most likely to pay off in those situations where opportunities to train are more scarce, the task is more demanding, and existing proficiency is unsatisfactory.

The 31Q experiment confirmed that substituting IVD in place of hands-on training can yield equivalent performance while reducing equipment costs. However, there are likely to be limitations to such substitution, and the 31Q experience points to them. Even though the IVD-trained students in the 31Q study were equally capable of accomplishing their tasks, they were slightly more likely to make certain procedural errors, and they expressed less satisfaction with training. We believe that these differences may arise from the extreme contrast in hands-on opportunity experienced in the two groups; the equipment-trained students enjoyed ample practice on real radio assemblages, whereas the IVD-trained group had only brief exposure to actual equipment. If true, this suggests that certain minimum levels of hands-on training may be required to ensure competency and self-confidence among trainees.

Thus, the results of both experiments indicate that IVD can be an effective element of training, and that there are conditions for using the technology wisely. However, given that proficiency was not dramatically affected in either application, and given the costs of acquiring IVD systems and developing supporting interactive courseware, the studies suggest that defense managers should give priority to those applications of IVD that can save training costs as part of a training resource mix. Further, we would argue that in applications designed to improve proficiency at added cost, the burden of proof should fall on the IVD proponent to show that improvement is needed and worth the cost.

ACKNOWLEDGMENTS

The authors acknowledge the support and encouragement provided by members of our sponsoring office, the Assistant Secretary of Defense for Force Management and Personnel. From the Defense Training and Performance Center, we received helpful advice and counsel from Dr. G. Thomas Sicilia, Director, Major Ron Tarr, and Mr. Tracy Mixon. We also acknowledge the assistance provided us by Dr. David Armor, formerly Principal Deputy Assistant Secretary for Force Management and Personnel, and by Mr. Gary Boycan of the Directorate of Training Policy.

This study would not have been possible without the support of the U.S. Army Signal Center, Fort Gordon, Georgia. In particular, we thank Dr. George Fry and Mr. Jim Frye of the Staff and Faculty Development Division, Directorate of Training and Doctrine, and the following individuals from the Transmissions Operations Department: Mr. Bob Shannon and Mr. Richard Hatcher of the 31M course, Mr. Robert Brock and Sergeant First Class James Brown of the 31Q course, and Mr. Dewey Plunket, Director. We are especially grateful to the numerous military and civilian course instructors in whose classrooms this research was performed, and whose cooperation was essential to the success of the evaluations.

We also acknowledge important contributions made by our RAND colleagues. A key contribution was provided by Neil Carey, who developed the performance measures used in our study in MOS 31Q and who cheerfully supervised this portion of the research. Throughout the entire effort, Debra Brown provided capable research assistance at Fort Gordon, while similar contributions were made at RAND by Marilyn Yokota and Ani Haigazian. We received valuable statistical advice from Robert Bell and James Press, and capable programming assistance from Jeff Garfinkle, Robert Young, and Scott Harrison. Richard Shavelson and Larry Hanser provided helpful technical reviews of an early draft. Melissa Huff and Kerrie Avery prepared the manuscript. The authors, of course, are solely responsible for any shortcomings.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS xi

FIGURES AND TABLES xv

Section

I. INTRODUCTION 1

II. BACKGROUND OF THE RESEARCH 4
- The Need for Training Technology 4
- Potential Benefits of Interactive Videodisc 5
- Origin of the Research 7
- Related Research on Training Technology 9
- Conclusions and Research Issues 20

III. SUPPLEMENTATION EXPERIMENT: THE 31M COURSE 25
- Overview 25
- Description of the 31M Course 26
- Experimental Design 30
- Performance Assessment 36
- Results 38

IV. SUBSTITUTION EXPERIMENT: THE 31Q COURSE 44
- Overview 44
- Description of the 31Q Course 45
- Experimental Design 47
- Performance Assessment 55
- Results 57

V. CONCLUSIONS 66
- Benefits of Experimentation 66
- Benefits of IVD for Supplementation and Substitution . 67
- Implications for IVD Development and Application 69

REFERENCES 73

FIGURES

3.1 Experimental and control classrooms 29
4.1 Arrangement of equipment in TROPO/LOS classrooms 48

TABLES

3.1 Characteristics of sample subjects in MOS 31M 31
3.2 Balancing variables, MOS 31M 33
3.3 Practical exercise training in the classroom 35
3.4 Regression results for task completion measures 40
3.5 Regression results for presence of REES errors 41
3.6 Predicted values from regression models 42
4.1 Characteristics of sample subjects in MOS 31Q 51
4.2 Balancing variables, MOS 31Q 52
4.3 Training received in experimental and control classrooms . 54
4.4 Regression results for performance of IF gain alignment . . 61
4.5 Regression results for performance of AGC alignment 61
4.6 Regression results for performance of squelch adjustment . 63
4.7 Predicted values from 31Q regression models 64

I. INTRODUCTION

Recent advances in defense modernization have brought complex new weapons and support systems into the inventories of the military services. However, improvements in military equipment present the defense training community with special challenges. To achieve intended improvements in capability, military personnel must be adequately trained to operate and maintain the complex new systems. Yet the training community also faces a number of countervailing pressures, including pinched training budgets, shortages of equipment available for training, and, as more training is shifted from training bases to the job, diminished means for assuring standardized, high-quality training.

Training organizations have begun to respond to these problems by employing new computer-based training devices and simulators. One device now receiving widespread interest is interactive videodisc (IVD) technology, which couples the interactive capability of the microcomputer with the high-fidelity visual capability of the laser videodisc. Using visual, interactive, and flexible presentation methods, such devices have the potential to simulate a variety of expensive equipment, place learners in dynamic problem-solving situations, and provide individualized training and feedback. However, such training technologies are themselves expensive and their training effectiveness has not been adequately assessed through rigorous evaluation. Equally important, there is currently no widely accepted or institutionalized method for determining the benefits of alternative military training devices.

This report presents the results of RAND research conducted to evaluate the effectiveness of an interactive microcomputer/laser videodisc (IVD) training system used to facilitate training in a variety of military occupational specialties in the Army, and increasingly, in the other services. The objectives of the study are to develop a methodology for assessing the benefits of innovative training technologies, to use the methodology to quantify the effectiveness of IVD training systems used in selected occupational specialties for advanced individual training, and to define beneficial future applications of similar training technologies.

Our approach was to apply principles of controlled experimentation to determine the effectiveness of an interactive videodisc system in two communications specialties at the U.S. Army Signal Center, Fort Gordon, Georgia. The methodology compared alternative approaches

to delivering training: one approach employed IVD, and both approaches were used to teach equivalent groups of trainees. Trainees were assigned to alternative conditions in a balanced, randomized design, using an established statistical model. They were subsequently compared on objective training- and job-related performance criteria. By isolating the cause of differences in performance to the method of training, the methodology provides precise experimental estimates of the effectiveness of alternative methods of training.

The applications were selected to represent two alternative uses of IVD: as a device used to *supplement* or augment existing hands-on training, and as a device used to *simulate* or replace at least some hands-on equipment training. Although the particular IVD applications examined are by no means representative of all such training in the Army, they are common applications. In the first case, IVD hardware and instructional material (courseware) are added to existing training resources; they increase training opportunity and they increase costs. In the second case, IVDs substitute for existing training resources; they can maintain existing training opportunity while decreasing costs.

The first application was examined in a study of Military Occupational Specialty (MOS) 31M, Multichannel Communications Equipment Operator, and the second approach was examined in a study of MOS 31Q, Tactical Satellite/Microwave Systems Operator. The principal findings of the 31M study showed that the use of IVD to supplement hands-on training yielded modest though statistically significant improvements in measures of proficiency, while increasing the costs of training. The 31Q study showed that groups of students trained under alternative regimens—one receiving hands-on training using only expensive equipment; the other receiving hands-on training using a mix of expensive equipment and lower cost interactive videodisc—are equally capable of performing the relevant tasks within established standards.

Our analyses further suggest training variables that may enhance or minimize IVD effectiveness in each of these types of application. Where IVD is to be used to supplement existing training, developers must attend carefully to the amount of existing practice opportunity, the difficulty of the task and the current level of proficiency, and the costs of adding the training technology. Where IVD is used to replace equipment, developers must identify the optimal mix of equipment and training technology that will permit sufficient practice on actual equipment, while still saving costs.

The remainder of the report describes in detail the background of the research, the methodology employed in the two studies, and spe-

cific results of the studies conducted in MOS 31M and MOS 31Q. Section II provides further information about the uses of IVD technology for military training and its development at the U.S. Army Signal Center, and it reviews relevant research on the training effectiveness of IVD and similar training technologies. Section III describes the research design and presents the results of the study conducted in MOS 31M. The 31Q study is described in Sec. IV. Section V summarizes the findings from both studies and discusses the implications of the research for DoD and Army policy regarding the development and implementation of IVD and similar training devices.

II. BACKGROUND OF THE RESEARCH

THE NEED FOR TRAINING TECHNOLOGY

The military training community expects an increased need for computer-based training devices and simulators to support training requirements. The constant introduction of new and advanced operational equipment has put pressure on the services' training systems, and the growing variety and complexity of new weapons and support systems have tended to raise skill requirements. In Army communications, for example, soldiers must be proficient with a growing variety of complex gear, and they must be able to sharpen their skills quickly as they change units and encounter new or unfamiliar equipment. The other branches and services face similar experiences, especially in the "high-tech" occupational specialties.

While the requirements for training have increased, training resources have not expanded in concert. The costs of many weapon and support systems limit their availability for training purposes at the training base or in field units. The time available for training at the training base has remained constant or been reduced. Furthermore, as the training burden is increasingly absorbed in units, problems of standardization increase: uniform instruction of consistent quality is hard to achieve.

In response to these challenges, military training departments are being urged to expand their use of various computer-based and visually oriented training technologies. Advisory groups such as the Defense Science Board[1] and the Army Science Board[2] have concluded that such training technologies as computer-aided instruction can improve greatly the readiness of the military force, while making training more efficient and effective. Both organizations recommend sizable new investments and an enhanced emphasis for training technology, simulators, and similar training devices.

The Army in particular is making a large investment in new training technology. Given the introduction of complex systems, the proliferation of paper-based training and technical materials, and a

---

[1]Defense Science Board, Office of the Under Secretary of Defense for Research & Engineering, *Report of the Defense Science Board 1982 Summer Study Panel on Training and Training Technology*, November 1982.

[2]Army Science Board, Office of the Assistant Secretary of the Army, *Final Report of the 1985 Summer Study on Training and Training Technology—Applications for AirLand Battle and Future Concepts*, December 1985.

felt lack of training resources, the Army identified requirements for "a generic information delivery system" that can provide "a more efficient, cost-effective method of delivering doctrinal, instructional, technical, operation, and maintenance materials to soldiers."[3] Interactive videodisc was identified as the required technology, embodied in a device called the Electronic Information Delivery System (EIDS), whose acquisition began in FY88. Initially, plans called for the acquisition of approximately 40,000 such systems, at an estimated cost of $200 million for hardware over an initial five years.[4]

POTENTIAL BENEFITS OF INTERACTIVE VIDEODISC

Capabilities of IVD Systems

The EIDS, produced by the MATROX Corporation of Canada, and its predecessor IVD systems[5] are systems for "communicative education and training" on computer and laser videodisc-related hardware that uses educational software (termed interactive courseware or "ICW"). Like similar methods of computer-based instruction, an IVD's powerful stand-alone microcomputer can provide individualized instruction, engage learners in dynamic problem-solving situations, and provide immediate feedback about performance. However, IVD goes beyond traditional computer-based instruction in its use of visual material. Current IVD units include a high-resolution color monitor tied to a laser videodisc containing up to 60,000 photographic frames. The video capability can provide a high-resolution representation of the target material in still-frame or motion sequences. This

[3]*EIDS Primer*, Department of the Army, Headquarters, U.S. Army Training Support Center, Fort Eustis, Virginia, n.d.

[4]The Army has since scaled back the scope of the acquisition because of budgetary pressures and concerns by the U.S. Army Training and Doctrine Command (TRADOC) that more time was needed to develop a comprehensive strategy for developing instructional material and fielding systems to units. For a description of current Army EIDS policy, see *Electronic Information Delivery Systems (EIDS) and Interactive Courseware (ICW) Implementation Plan*, Department of the Army, Headquarters, United States Army Training and Doctrine Command, Fort Monroe, Virginia, January 1988. For a description of how Army schools and TRADOC now select IVD projects for development, see J. D. Winkler, "Army Applications of Interactive Videodisc Technology," The RAND Corporation, forthcoming.

[5]An "interim" system consisting of a Sony SMC-70 microcomputer, PVM-120Q monitor, LDP-10000A videodisc player, and floppy disk drives has been in common use throughout the Army.

capability allows IVD to act as a two-dimensional or so-called "generic simulator" for a variety of new equipment.[6]

Capacity for Simulation

IVD represents a significant advance in training technology. Although the services have employed various forms of computer-based instruction since the 1960s,[7] a principal advantage of IVD is its facility for simulation. Users of IVD can view equipment and, with use of a peripheral device such as a light pen, can simulate tasks and procedures such as adjusting controls and inserting cables. The photographs can be used in motion sequences, for example, as the movement of a meter or the firing of a missile. There is also an audio track for simulating sounds associated with equipment, as for example, in the sound of a generator running after it has been properly installed. With the control provided by the computer, the system can present visual information, accept responses, record errors, branch to remediate these errors, and, in general, offer individualized and interactive simulations.

Training and Cost Benefits

Advocates argue that IVD technology has important training benefits.[8] IVD is believed to improve proficiency with job skills. It is also believed to improve classroom productivity by improving the amount and quality of training time when equipment is scarce, because there is less "slack time" while trainees wait for opportunities to train on equipment. In the training school environment, this improvement in productivity could translate into an increased ability to process more trainees faster during a mobilization. There may also be cost savings associated with the use of IVD.[9] IVD hardware is usually less expensive than the equipment that it may replace. If training time is shared between actual equipment and IVD, then cost savings should also result from less wear and tear on actual equipment.

---

[6]J. W. Clark, "Videodisc Training at Fort Gordon: A Practical Application," *The Videodisc Monitor*, July 1985, p. 10.

[7]J. Orlansky and J. String, "Computer-based Instruction for Military Training," *Defense Management Journal*, 1981, pp. 46–54.

[8]See, for example, D. Best, M. Caldwell, and L. Harrelson, "Video Disc Training," *Army Trainer*, Vol. 2, No. 3, 1983, pp. 7–9.

[9]Each unit of EIDS hardware costs $4000–$8000; when bought off the shelf, the cost of interim Sony systems has been approximately $4000–$6000. These figures do not include the costs of courseware development, which can be considerable.

Advocates of IVD envision applications in active duty and reserve units, for example, as part of equipment-specific on-the-job training. If this were true, some pressure on the schoolhouse could be relieved. IVD is also believed to be useful in sustainment and refresher training, particularly in cases when training is intermittent or skills can decay, as when servicemen are assigned to units with unfamiliar equipment. Finally, given the distributed nature of much of reserve training, IVD could be a useful adjunct to reserve schools and units—one that ensures that reserve trainees receive standardized instruction and up-to-date training.

ORIGIN OF THE RESEARCH

IVD Use at the U.S. Army Signal Center

The U.S. Army Signal Center at Fort Gordon, Georgia, has been a leader in developing and implementing IVD technology in military occupational training. The Signal Center trains approximately 33,000 individuals per year in technical communications specialties. It pioneered the use of low-cost, off-the-shelf IVD equipment to facilitate training. This use evolved into a methodology for configuring hardware and developing in-house instructional courseware into integrated interactive training systems. The systems are designed to expose trainees to more types of operational equipment and procedures than is possible through regular classroom instruction. They also provide interactive self-paced instruction and testing and produce a record of each trainee's responses during every session.

The Signal Center has produced most of the IVD training material available in the Army.[10] By 1987, the Signal Center had completed approximately 37 videodiscs, encompassing approximately 750 hours of instruction, for use in six different communications-electronics occupational specialties.[11] Another 36 videodiscs, covering an additional 1300 hours of instruction in seven specialties, were at various stages of development. Some 35–40 people were actively involved in the development of these videodiscs.[12] As their experience grew, the

---

[10]*Supplementary Interactive Courseware User's Guide (ICUG) Bulletin*, Department of the Army, Headquarters, United States Army Training Support Center, Fort Eustis, Virginia, August 1987.

[11]"Basis of Issue Plan for Electronic Information Delivery System (EIDS)," Department of the Army, Headquarters, U.S. Army Signal Center and Fort Gordon, Fort Gordon, Georgia, 1987.

[12]Clark, July 1985.

IVD developers, housed in the Training Technology Branch, Staff and Faculty Division of the Directorate of Training and Doctrine, began offering training workshops in videodisc development and production to other Army schools and representatives from the other services.[13] In 1985, the Signal Center was named by the Army as the "combat developer" for the EIDS system—the organization responsible for ensuring that the EIDS system was fielded to meet the users' needs.

Need for Evaluation

Like other IVD advocates, Signal Center IVD developers believe that IVD systems can significantly increase trainee proficiency, reduce hands-on training requirements, and offer training support in a broad range of specialties. They also believe that IVD systems have potential for refresher and on-the-job training in field units. As the official Army proponent for the EIDS system, the Signal Center desired to assess the effects and identify the most productive applications of IVD technology. Because EIDS had once been designated as the DoD videodisc standard by the Defense Visual Information Standardization Committee,[14] and because the other military services were interested in IVD-based training technologies, the Office of the Secretary of Defense (OSD) also seeks to establish the training effectiveness of IVD and ensure that videodisc technology is wisely implemented. If IVD training could be shown to be advantageous, and if research could distinguish potential high-payoff applications, demonstration of these benefits could affect decisions by the Army and the other services about purchasing EIDS or similar IVD hardware, developing EIDS/IVD courseware, and allocating such systems across various specialties and environments. However, the systematic data needed to assess these potential benefits have been lacking to date. Moreover, there is no commonly agreed-upon method for assessing the training effectiveness of new technologies such as IVD or EIDS.

In 1985, the Signal Center expressed an interest in sponsoring systematic research to evaluate the effectiveness of interactive videodisc training. The Signal Center and TRADOC asked the Defense Training and Performance Data Center (TPDC) to provide analytical support for this research. TPDC in turn asked The RAND Corporation to design, perform, and analyze the research. This report

---

[13]J. W. Clark, "Videodisc Training Workshop: Fort Gordon, Georgia," *The Videodisc Monitor*, May 1985, p. 10.

[14]P. F. Gorman, "Educational Technology: Yesterday, Today and Tomorrow," *Military Review*, Vol. 66, No. 12, December 1986, pp. 4–11.

represents the results of the two-year research effort that emerged from these requests.

RELATED RESEARCH ON TRAINING TECHNOLOGY

To determine the appropriate methods for our research, and to identify appropriate courses and tasks in each stage of our investigation, we examined a range of IVD applications at the Signal Center and in the Army. We also sought insights in the research literature. Broadly, the literature covers the following major issues, in increasing order of specificity:

- Computer-based instruction (CBI) and related interactive technologies, including numerous evaluation studies and reviews of the field
- Simulation, encompassing evaluations of the benefits of specific devices and computer programs
- Interactive videodisc technology, including descriptions of the technology's promise and some empirical studies of its benefits.

Each of these issues comprises a category of the literature, and each has its own subcategories. Each includes studies conducted in military training or civilian educational contexts. Among the empirical studies found in each area, laboratory research predominates, but a smaller number of field studies are also found.

The remainder of this section draws on the literature that we judged relevant. Our primary criterion was that it could ultimately inform, through its implications for our research, the design of policies for ensuring that IVD technology, if acquired, was used most productively.

The general literature on the effects of computer-based instruction is informative for its general findings and for its discussion of issues related to evaluation of CBI as a training medium. We begin first with an overview of this literature. Especially relevant are studies of IVD in military training, particularly empirical analyses that provide quantitative estimates of its training effectiveness. By "training effectiveness analyses" we follow conventional definitions as offered by the U.S. Army Training and Doctrine Command: research in which the goal is "an assessment of proficiency in an effort to determine the effectiveness of training."[15] In this context, the research should

[15]Department of the Army, U.S. Army Training and Doctrine Command, *Training Effectiveness Analysis: A Process in Evolution*, TRASANA Pamphlet 350-4, U.S. Army

assess the competence or proficiency at tasks trained using IVD. Thus, we discount certain studies, such as those examining student or instructor "acceptance" of IVD technology, which do not contain measures of student performance.[16]

For similar reasons, research that does not contain measures of performance relevant to military training is not reviewed. We also identified a small number of studies that evaluate IVD programs used in civilian contexts, for example, to teach physiology, chemistry, or arithmetic.[17] The outcome measures examined in most of these studies are generally relevant to concerns of education, but less applicable to the training of military tasks. Commonly these studies examine student achievement as measured in written achievement tests, attitudes toward the educational technology (e.g., measures of "acceptance"), or social effects (e.g., on classroom interaction).

Research on Computer-Based Instruction

The literature examining the effectiveness of computer-based instruction and collateral issues (e.g., effects of interactivity on learning) is large; it is not our intention to review its findings in detail here.[18] The literature is instructive, however, in two important respects. First, the accumulated research findings suggest the nature

---

TRADOC Systems Analysis Activity, White Sands Missile Range, New Mexico, August 1985.

[16]See, for example, D. T. Manning, "Student Acceptance of Videodisk-Based Programs for Paramedical Training," *T.H.E. Journal*, Vol. 11, No. 3, 1983, pp. 105–108.

[17]See, for example, Charles E. Branch, B. R. Ledford, B. T. Robertson, and L. Robison, "The Validation of an Interactive Videodisc as an Alternative to Traditional Teaching Techniques: Auscultation of the Heart," *Educational Technology*, March 1987, pp. 16–22; Barbara Gross Davis, *Nebraska Videodisc Science Laboratory Simulations* (Executive Summary), The Annenberg/CPB Project, University of Nebraska, March 1985; Ted Hasselbring et al., "An Evaluation of a Level-One Instructional Videodisc Program," *Journal of Educational Technology Systems*, Vol. 16, No. 2, 1987–88, pp. 151–169.

[18]Readers interested in this issue should consult the following reviews: Henry J. Becker, *The Impact of Computer Use on Children's Learning: What Research Has Shown and What It Has Not*, Center for Research on Elementary and Middle Schools, Johns Hopkins University, n.d.; Gerald W. Bracey, "Computers in Education: What the Research Shows," *Electronic Learning*, November-December 1982, pp. 51–55; J. Edwards, S. Norton, S. Taylor, et al., "How Effective is CAI? A Review of the Research," *Educational Leadership*, November 1975, pp. 147–153; Richard Niemic and H. J. Walberg, "Comparative Effects of Computer-Assisted Instruction: A Synthesis of Reviews," *Journal of Educational Computing Research*, Vol. 3, No. 1, 1987, pp. 19–37; J. Orlansky, "Effectiveness of CAI: A Different Finding," *Electronic Learning*, Vol. 3, No. 1, September 1983, pp. 58–60; and John F. Vinsonhaler and R. K. Bass, "A Summary of Ten Major Studies on CAI Drill and Practice," *Educational Technology*, July 1972, pp. 29–37.

and magnitude of the effects one should expect from CBI as an instructional medium. Insofar as IVD technology depends on the interactive capability provided by its microcomputer, the findings should suggest potential effects of IVD. We shall discuss both the effects on achievement and the effects on instructional time. Second, the literature offers insight on the methodologies and their limitations for evaluating the benefits of innovative instructional technologies.

Effects on Achievement. A major concern of many studies is the effect of the medium of instruction (CBI) on student learning. Because many research studies have addressed this issue, their results have been synthesized using *meta-analysis*, a technique for combining the effects of independent research studies.[19] This technique summarizes disparate research results using a common statistical metric, termed *effect size*. Effect size is calculated using sample means and standard deviations as reported in each study, or it can be calculated from covariance-adjusted means or other statistics such as the *t*-test. Effect sizes are often calculated as the difference between the outcome means of the experimental and control groups, divided by the standard deviation of the control group.[20] The difference between the groups is stated as an improvement or decrease in units of standard deviations.[21] It may also be transformed statistically into percentile scores for each group.

Kulik and his colleagues have conducted several meta-analyses of the effects of CBI in various educational settings, including elementary, secondary school, and college.[22] The studies scrutinized through meta-analyses are only those conducted in actual classrooms, comparing groups of computer-taught and conventionally taught stu-

---

[19]G. V. Glass, B. McGaw, and M. L. Smith, *Meta-analysis in Social Research*, Sage Publications, Beverly Hills, California, 1981.

[20]An alternative method, which we prefer, is to divide the difference of the means by the pooled standard deviation of the experimental and control groups. See J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, revised edition, Academic Press, New York, 1977.

[21]Generally, in comparing means of two groups, differences less than or equal to 0.20 standard deviation are regarded as "small," whereas those greater than or equal to 0.80 standard deviation are regarded as "large." Values intermediate to these are regarded as "medium" (Cohen, 1977, pp. 23ff).

[22]Robert L. Bangert-Drowns, James A. Kulik, and Chen-Lin C. Kulik, "Effectiveness of Computer-Based Education in Secondary Schools," *Journal of Computer-Based Instruction*, Vol. 12, No. 3, 1985, pp. 59–68; James A. Kulik, Chen-Lin C. Kulik, and Peter A. Cohen, "Effectiveness of Computer-Based College Teaching: A Meta-analysis of Findings," *Review of Educational Research*, Vol. 50, No. 4, Winter 1980, pp. 525–544; J. A. Kulik, Chen-Lin C. Kulik, and R. L. Bangert-Drowns, "Effectiveness of Computer-Based Education in Elementary Schools," *Computers in Human Behavior*, Vol. 1, 1985, pp. 59–74.

dents, and free of crippling methodological flaws. The general finding in these studies is that computer-based education had positive effects on student achievement. The magnitude of the effect differs in the populations, however. Among secondary school students in the average study, test scores rose by approximately .40 standard deviation for programs of computer-assisted instruction and computer-managed instruction.[23] Among elementary students, the average improvement was .47 standard deviation,[24] whereas among college students, the average improvement was smallest (.25 standard deviation).[25]

Which of these studies should be regarded as most relevant to military training? Fortunately, the issue was addressed in another meta-analysis that included studies of CBI in military training that fulfilled the authors' methodological criteria for inclusion.[26] Based on 24 controlled studies on adult education, including 10 studies of military training, the average improvement in learning (based on examination scores), was .42 standard deviation. In percentile scores, this suggests that CBI raised the performance of the typical student from the 50th to the 66th percentile. In the metric of effect sizes, the magnitude of difference is regarded as "moderate."[27] We thus regard it as suggestive of the size of the effect that could be found in comparative evaluations of interactive videodisc technology.

Effects on Instructional Time. A second generalization emerging from the literature is that CBI can reduce the amount of time needed to train or educate the learner. This finding emerges both from the meta-analyses of CBI studies containing measures of instructional time[28] and from conventional reviews of the effects of CBI in military training.[29] Indeed, based on the results of a conventional literature review, researchers at the Institute for Defense Analysis have concluded that the evidence for improved achievement from CBI is weak, and that the principal benefit of CBI in military training is that it "saves students time in attaining the required minimum levels

---

[23]Bangert-Drowns, Kulik, and Kulik, 1985.

[24]Kulik, Kulik, and Bangert-Drowns, 1985.

[25]Kulik, Kulik, and Cohen, 1980.

[26]Chen-Lin C. Kulik, J. A. Kulik, and B. J. Schwalb, "The Effectiveness of Computer-Based Adult Education: A Meta-analysis," *Journal of Educational Computing Research*, 1986.

[27]Kulik, Kulik, and Schwalb, 1986.

[28]Kulik, Kulik, and Cohen, 1980; Kulik, Kulik, and Schwalb, 1986.

[29]J. Orlansky and J. String, "Cost-Effectiveness of Computer-Based Instruction in Military Training," Institute for Defense Analysis, IDA Paper P-1375, April 1979.

of knowledge and skills without a loss of student achievement."[30] The median time savings in 19 studies was on the order of 30 percent. This finding suggests that CBI or related technologies could be used to shorten and thus decrease the costs of training, assuming of course that training course duration is free to vary.

Methodological Criticisms. The literature on CBI also contains an extended critical discussion of the disadvantages of comparative evaluation in assessing the effectiveness of CBI.[31] A principal claim of critics is that the research comparing CBI-delivered instruction to that delivered by traditional means is "confounded" The argument states that the "treatment condition" in most evaluations consists of the medium of instruction (CBI), plus uncontrolled effects arising from different instructional content, teaching methods, or novelty in the alternative classrooms. Failure to match instructional content means not only that instructional materials may differ, but that the amount of instruction or practice can vary.[32] Thus when CBI is compared with some other medium, usually teachers, the differences observed may be due to factors other than the CBI itself.[33]

We do not entirely accept these criticisms, for reasons that will be discussed fully in the conclusion to this section. Although research in specific academic traditions may seek to differentiate the causal effects of technological media and other elements of instruction, policy research seeks to identify the primary effects of alternative "packages" of training resources.[34] Policymakers need to know the benefits that can be expected from various training approaches in order to guide decisions about major expenditures of public funds.

---

[30]Orlansky, 1983, p. 58.

[31]R. E. Clark, "Confounding in Educational Computing Research," *Journal of Educational Computing Research*, Vol. 1, No. 2, 1985, pp. 137–148; P. Hagler and J. Knowlton, "Invalid Implicit Assumption in CBI Comparison Research," *Journal of Computer-Based Instruction*, Vol. 14, No. 3, September 1987, pp. 84–88; G. Salomon and H. Gardner, "The Computer as Educator: Lessons From Television Research," *Educational Researcher*, Vol. 15, No. 1, January 1986, pp. 13–19; Theodore M. Shlechter, *An Examination of the Research Evidence for Computer-Based Instruction in Military Training*, U.S. Army Research Institute for the Behavioral and Social Sciences, ARI Field Unit, Fort Knox, Kentucky, August 1986.

[32]Schlecter, 1986.

[33]Some critics of evaluations further argue that such outcome comparison studies are pointless and that comparative analyses of media effectiveness should be abandoned in favor of theory-based research or "holistic" descriptive studies; these latter examine, for example, the attributes and capacities of computers for delivering instruction or individual differences in learning approaches using CBI. See R. E. Clark, 1985, p. 141; Salomon and Gardner, 1986, p. 16.

[34]R. J. Shavelson et al., *Evaluating Student Outcomes from Telecourse Instruction: A Feasibility Study*, The RAND Corporation, R-3422-CPB, May 1986.

Systematic evaluations, employing objective measures of job performance to quantify the expected benefits from altering existing training, remain the most appropriate method for addressing the policy question. Furthermore, although we agree that instructional content and strategy should be matched as closely as possible in any comparative evaluation of IVD, control over content should be easier to accomplish in studies of military training, where specific job-related tasks, competencies, and training approaches are defined within an established program of instruction. Thus it is unlikely that alternative groups would receive fundamentally different training. Nevertheless, where the amount of instruction or practice may vary, such differences should be explicitly monitored as part of the research design.

Research on Training Applications of Interactive Videodisc

We now turn to the literature assessing the specific benefits of interactive videodisc for providing military training. Our review is confined to studies examining IVD use in actual training courses; the studies that we located have all occurred within the Army. One might hope that with the imminent implementation of EIDS technology, systematic assessments of IVD effectiveness might be in hand to guide policy about how to use the systems most productively. Unfortunately, few such studies have been performed, and of these, most suffer from important methodological limitations. We discuss three sets of studies on IVD in military training: studies of IVD in Army communications training, studies of IVD in Army medical training, and other Army studies of IVD effectiveness.

IVD in Army Communications Training. Several prior studies of the training effectiveness of IVD have been conducted at the Signal School. The first examined the use of an IVD system to provide hands-on training in MOS 26Y, Satellite Communications Systems Repairer.[35] The purpose of the study was to determine if IVD could provide substitute training for a more expensive ground satellite terminal. The Signal School was concerned about a shortage of available equipment and the cost of maintaining the equipment, which was not designed to be "powered up" and "powered down" repeatedly. An IVD system was developed to increase training opportunity and maintain

[35]W. D. Ketner, "The Videodisc/Microprocessor for Training," *Training and Development Journal*, May 1981, pp. 151–153; J. I. Young and D. T. Tosti, *Equipment-Independent Training Program*, U.S. Army Training and Doctrine Command, Training Developments Institute, Report TDI-TR-3-81, December 1981.

or improve existing levels of proficiency, while reducing wear and tear on equipment.

An experiment was conducted using one lesson—a three-hour practical exercise of alignment procedures of a satellite communications ground terminal (AN/FCC-98). All students in the appropriate segment of the course were assigned at random to one of two conditions. An experimental group of 27 students received training using only the IVD system, while a comparison group of 24 students practiced on actual equipment. The groups were compared on a hands-on test of performance on actual equipment (rated "Go" or "No Go") and on written examinations.

The authors of the study report that the groups did not differ on any of the measures at conventional levels of statistical significance. They conclude that "both forms of practice are equally effective" for training.[36] Unfortunately, this conclusion is not clearly implied by the data. Performance on both the hands-on test and on the written examination favored the group using equipment ($t = -1.43$ and $t = -1.11$, respectively). Small sample size may have been responsible for the lack of statistical significance in the difference between the groups.

A second experiment, also conducted in MOS 26Y, examined the effectiveness of an IVD system for training soldiers to program an expensive multiplexer system (AN/GSC-24) in short supply in the course (four were available in a class of approximately 65 students).[37] Students in one class were randomly assigned to one of two conditions. An experimental group, consisting of 28 students, used only an IVD system during a three-day laboratory exercise, while a control group of 31 students practiced on the actual equipment. Thus, IVD again substituted for equipment training. The experimenters measured the amount of practice received and subsequent performance on the portion of the hands-on test devoted to programming the multiplexer.

The results of the study showed that the group that practiced on the IVD system received considerably more training opportunity within the laboratory period.[38] Nonetheless, the groups were statistically indistinguishable on two performance measures—hands-on

---

[36]Young and Tosti, 1981, p. I-5.

[37]G. L. Wilkinson, *An Evaluation of Equipment-Independent Maintenance Training By Means of a Microprocessor-Controlled Videodisc Delivery System*, Battelle Memorial Institute, Report TDI-TR-83-1, March 1983.

[38]The experimenters did not control for training opportunity. The control group had more training stations available than did the experimental group.

test score and time to complete the test. We again caution, however, that the conclusion that the IVD system in this study was "an effective alternative" or "more efficient" than hands-on training[39] is weakened by the small sample size. The hands-on group performed better than the IVD-trained group, but because the sample size in the study was small, the differences may not have achieved statistical significance. Moreover, this study raises other methodological concerns. The groups differed in important ways that were confounded with treatment. Despite randomization, the data analysis indicated statistically significant group differences in education and skill at using multiplexers earlier in the course, in addition to amount of practice, yet the analysis did not control for those differences. The measurement properties of the performance measure were also undefined. We thus regard these results as only suggestive.[40]

A later study examined a different application of IVD training in MOS 72G, Automatic Data Telecommunications Center Operator.[41] The purpose of this study was to examine the effectiveness of using IVD to supplement hands-on training on scarce equipment. The DCT-9000 is a complex data communications terminal with associated components. Only one system was available to teach a class of approximately 18 students, and each student received less than two hours of practice during a one-week course module. The cost of the DCT-9000 (approximately $300,000 in 1984) precluded the acquisition of additional equipment; thus, IVD was identified as a lower-cost method of providing additional practice.

This study, unlike the studies described above, compared the effects of hands-on training on equipment with training provided by equipment *and* IVD. An experimental group of 76 soldiers used two IVD systems in addition to the DCT-9000, while a comparison group of 74 soldiers were trained using equipment only. The students were *not* assigned to groups at random or trained concurrently, however. The comparison group participated during a "baseline" interval; the IVD systems were then introduced, and subsequent trainees constituted the experimental group. Although the data showed no difference between the groups in population characteristics, the effects of using such a design are nonetheless subject to alternative interpreta-

---

[39]Wilkinson, 1983, p. I-5.

[40]Interestingly, the experimental group was significantly more negative toward the training received, as indicated on attitudinal measures. Their comments indicated that they were most unhappy at having received no hands-on training.

[41]C. D. Vernon, *Evaluation of Interactive Video Disc System for Training the Operation of the DCT-9000 in the MOS 72G Course*, U.S. Army Communicative Technology Office, Report TR-84-6, October 1984.

tions, including historical differences and greater likelihood of so-called "Hawthorne effects."[42]

The experimenters monitored the training provided to each group and subsequently assessed the performance of the soldiers on hands-on performance tests on consecutive days. The results showed that the experimental group received more than double the practice time of the comparison group, and they performed significantly better on the initial hands-on test (an improvement of nearly 7 percent). The groups scored equally well on the retest. The scores on the retest were extremely high for both groups (mean of 96 percent); thus, a "ceiling effect" in the outcome measure may have precluded group differences. The results of this study favor IVD training, although the results also suggest that the equipment training may have been sufficient to achieve proficiency within the allotted time.

A final study in a Signal-related specialty was conducted at the Army Intelligence School, Fort Devens, Massachusetts, in MOS 33S, Signal Intelligence/Electronic Warfare Systems Repairer.[43] This study also examined the effectiveness of an IVD system as a supplement to hands-on training on actual equipment. The research design was similar to that used in MOS 72G and is subject to the same limitations. A baseline group of 51 students was trained to operate and troubleshoot using the radio receiver RACAL R-2174(P). Subsequently, 48 students received similar training using a combination of IVD and equipment. The course segment was two weeks. Practice time was recorded, and all trainees received a hands-on test of troubleshooting and a written test.

The introduction of IVD provided the experimental group with an increase in training time of 28 percent, compared with the baseline group. Although mean values obtained on the performance measures favored the IVD-trained group (in both hands-on and written tests), the size of the differences did not achieve conventional levels of statistical significance. Unfortunately, the study may have been victimized by its small sample size; the improvement in hands-on performance amounted to 5 percent ($t = -1.85$). Had the magnitude of the group difference held up in a larger sample, the difference would have been statistically significant, although still not very large.

---

[42]See D. T. Campbell and J. C. Stanley, *Experimental and Quasi-experimental Designs for Research*, Rand-McNally Publishing Co., Chicago, 1963, for a description of the threats to validity encountered by such nonexperimental designs.

[43]G. L. Wilkinson, *Evaluation of the Effectiveness and Potential Application of the Interactive Videodisc Instructional Delivery System within the Training of SIGINT/EW Systems Repairers*, Battelle Memorial Institute, Report TRS-85-1, March 1985.

IVD in Army Medical Training. Other than the studies reviewed above, the closest analog to a program of research on the training effectiveness of IVD was conducted at the U.S. Army Academy of Health Sciences, where researchers examined the value of IVD for teaching combat medics (MOS 91A) to give intramuscular injections. Generally, the studies compared one or more versions of IVD-based training with traditional methods of instruction, which involved preparing and administering an injection to another student in the class.

An initial study compared a group of students taught by conventional methods with a group in which IVD was used by the instructor to "enhance" (i.e., supplement) the existing training.[44] Students were assigned at random to either the control condition (N = 42) or the experimental IVD condition (N = 28). The design of the study was unusual in that the instructors could terminate the instruction when they felt comfortable with the progress of the students.

Proficiency tests were given two days and 17 days after training. The latter test was an unannounced "surprise test" designed to compare the groups on their retention of the tasks. The results of the study showed that the instructors in the IVD-trained group terminated their training earlier (an improvement of 43 percent in training time compared with the control group), and the experimental students were significantly more likely to pass the "surprise" test (76 percent in the experimental condition compared with 59 percent in the control condition). However, the groups were equally likely to pass the first test.

A later study compared a group of students taught by traditional methods with two groups using IVD in place of traditional methods.[45] This application of IVD, then, most resembles the "substitution" experiments in MOS 26Y described above. Each group contained 84 students, who were randomly assigned. Instructors were also randomly assigned to classrooms and sensitized regarding possible "Hawthorne effects" Students were tested for proficiency two days after training and again in an unscheduled test after 15 days. Assessors "blind" to the experimental condition of the trainees judged their success or failure at administering an intramuscular injection. The results of the study showed once again that savings in training

---

[44]D. G. Ebner, D. T. Manning, F. R. Brooks, et al., "Videodiscs Can Improve Instructional Efficiency," *Instructional Innovator*, Vol. 29, No. 6, 1984, pp. 26–28.

[45]P. M. Balson, D. T. Manning, D. G. Ebner, et al., "Instructor-Controlled Versus Student-Controlled Training in a Videodisc-Based Paramedical Program," *Journal of Educational Technology Systems*, Vol. 13, No. 2, 1984–85, pp. 123–130.

time were achieved in the IVD-trained groups, whereas the number of students who failed the task of administering an intramuscular injection did not differ between the groups. Unfortunately, the article does not report the proportion of students passing in each group. If the passing rates were uniformly high, the groups might not differ on that basis alone.

The final study conducted at the Academy of Health Sciences compared alternative approaches to using IVD. IVD was used in all groups to enhance existing instruction.[46] Trainees (N = 246) were randomly assigned. The control group received one exposure to IVD in an audiovisual demonstration and the experimental groups received "limited" or "full" access to the IVD training materials. Outcome measures again consisted of time spent teaching the task "to proficiency" and a hands-on test of administering an intramuscular injection.

As in the earlier studies, the authors report a savings of training time and improvements in proficiency in the experimental groups compared with the control group. The account of the research again raises questions, however. Given the definition of the control group, the meaning of the differences is not straightforward; they appear to represent the results of varying the amount of IVD exposure. Moreover, the only statistics given in the report assert "7 to 8% superiority" in the experimental groups; no other information about the distribution of outcomes or test statistics is provided. Thus, we are unsure of how to interpret these findings.

Other Army IVD Training Studies. We found few other studies examining the training effectiveness of interactive videodisc technology. One study examined the effectiveness of IVD for delivering training extension course (TEC) lessons.[47] Two groups, each containing approximately 100 soldiers, viewed TEC lessons appropriate to their MOS on either a prototype IVD player or in super-8mm using a Bessler Cue/See. Members of the comparison group received no training at all. Soldiers were then administered hands-on tests of the TEC material. Not surprisingly, given their lack of training, the members of the control group performed more poorly than members of either experimental group. The results suggest that practice by any means is preferable to no practice.

---

[46]P. M. Balson, D. G. Ebner, J. V. Mahoney, et al., "Videodisc Instructional Strategies: Simple May Be Superior to Complex," *Journal of Educational Technology Systems*, Vol. 14, No. 4, 1985–86, pp. 273–281.

[47]J. E. Holmgren, F. N. Dyer, R. E. Hilligoss, and F. H. Heller, "Effectiveness of Army Training Extension Course Lessons on Videodisk," *Journal of Educational Technology Systems*, Vol. 8, No. 3, 1979–80, pp. 263–274.

Another study at the U.S. Army Armor Center examined performance at tank gunnery in an evaluation of a videodisc-based simulator, the "VIGS."[48] A group of 20 soldiers who received conventional one-station unit training were compared with 20 soldiers receiving VIGS training as a supplement. This preliminary account observes that the additional VIGS training resulted in faster engagement times and fewer procedural errors. However, no statistical analyses had been performed at the time the document was prepared.

CONCLUSIONS AND RESEARCH ISSUES

Based on the above review, we draw the following conclusions regarding approaches to evaluating IVD technology. These conclusions encompass the methodological and substantive issues we consider most important for designing research to assess IVD training effectiveness.

Appropriate Research Methods

Despite criticism in the literature on CBI, we remain convinced that systematic comparison is the appropriate method for evaluating the effectiveness of innovative training technologies such as IVD. Theory-based and descriptive research studies may be appropriate for academic research on improving curriculum or the design of computer courseware, but such research cannot offer much help to policymakers who must decide how much to spend and how to deploy innovative training technologies. The Army, in particular, planned to acquire up to 40,000 EIDS systems at a cost for hardware of approximately $200 million. Managing an investment of this magnitude and deciding on further investments require concrete information on expected benefits. Such guidance can come only from empirical demonstrations of benefits received under alternative conditions of use.

Although in general such demonstrations should attempt to specify particular conditions that maximize effectiveness, we do not agree with some of the CBI literature's strictures against "confounding" technological media with other elements of instruction. Policy guidance for using IVD does not require that the effects of innovative training strategies be reduced strictly to the training medium while holding constant all other elements of training. Rather, policy guidance requires knowledge of the expected benefits of the technology, on

[48]John A. Boldovici, "VIGS Evaluation," Memorandum, U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Knox, Kentucky, June 1986.

average, as used in practice in its various applications and settings. The Congressional Office of Technology Assessment draws an analogous distinction in discussing the evaluation of innovative medical technologies.[49] It defines "effectiveness" as "the benefit of technology under average conditions of use" in the typical clinical setting.[50] For instance, even though a medical treatment might be used by different doctors, on different patients, in different settings, for dissimilar symptoms, one still needs to know the average expected benefit before making major social investments. The consensus in medicine is that the ideal information derives from a clinical trial that compares the medical innovation to an established treatment. We believe that similar reasoning applies to the evaluation of training technologies (e.g., IVD), where the "clinical trial" examines the effects in various training settings.

We conclude not only that the training effectiveness of IVD should be evaluated using comparative methods, but that it should use the strongest possible designs. Unfortunately, much of the previous research on IVD training effectiveness is limited by problems with the research design or statistical analysis that diminish our confidence in the findings. We note, in particular, three methodological problems.

First, many studies have used designs with insufficient statistical power to detect effects between groups. This weakness is especially problematic in studies that examine IVD use as a substitute for hands-on training. In studies that posit "no difference" or equivalence of outcomes, sample size must be sufficiently large to ensure confidence in lack of difference as a conclusion. Second, quite a few studies, particularly those concerned with IVD supplementation, suffer from non-equivalence of "treatment" and "comparison" groups, through failure to randomize trainees to treatment or to compare training methods concurrently. Randomized experiments are the strongest possible methods for establishing causal relationships between independent variables (e.g., training method) and outcome variables (e.g., job proficiency). Third, the studies that we reviewed often used statistical analyses that failed to control for potentially confounding effects, such as differences in demographic background. Such methodological flaws should be corrected in future research on IVD effectiveness.

---

[49]Office of Technology Assessment, U.S. Congress, *Strategies for Medical Technology Assessment*, U.S. Government Printing Office, Washington, D.C., 1982.

[50]Office of Technology Assessment, 1982, p. 33.

Potential IVD Applications

How should IVD be examined in a randomized experiment? The literature on IVD training effectiveness shows that supplementation and substitution are principal uses of IVD. The rationale for acquiring IVD in each of these applications is quite different. In the case of supplementation, the amount of available hands-on training is deemed insufficient, primarily because the amount of available equipment is perceived as inadequate and additional equipment is difficult to obtain.[51] In these situations, the principal effect of introducing IVD is to increase the amount of practice. The additional practice is intended to increase training "productivity" (i.e., time spent practicing tasks in the classroom); it is further expected to improve subsequent task proficiency. Clearly, however, the addition of IVD resources also increases the costs of training.

In the case of substitution, IVD is acquired to replace training that has been provided by some other means. Existing training may be seen as too costly, difficult, or dangerous, or otherwise unsatisfactory.[52] In this situation, IVD is intended to provide equivalent training at less cost (e.g., in equipment acquisition or maintenance, training time, or hazard to trainees). Although IVD is substituted, it may be considered less desirable than the hands-on training it may replace. However, proficiency is expected to remain at least equivalent to that of the "traditional" methods of training being replaced.

Based on discussions with Army IVD developers and observations of many Army IVD programs, our impression is that these two approaches are the most common ways in which IVD is used, and supplementation is a more common application than substitution. These are the major types of IVD training applications that should be evaluated. Further, as each represents an alternative to existing training, each should be compared to the current approach in use. Where IVD is acquired to augment hands-on training (at increased costs), the effects of the extra IVD practice should be quantified relative to the practice provided by hands-on training. Where IVD has been acquired to substitute for hands-on training, the effects of the resources that are substituted should be compared to the effects of the training resources that are replaced.

---

[51]Examples in the literature were the cases of the DCT-9000 and the RACAL R-2174(P) in MOSs 72G and 33S, respectively.

[52]Examples in the literature were the satellite terminal repair task in MOS 26Y and the task of intramuscular injection in MOS 91A.

Potential Effectiveness of IVD Applications

We have learned from the literature that IVD applications can have three primary effects: improving task proficiency, saving costs of resources, and reducing training time. The literature on CBI effectiveness, based on meta-analytic findings, suggests that the expected effects of IVD training on task proficiency should be positive, although they may be modest in size. Meta-analyses of CBI effects show improvements in the range of one-quarter to one-half of a standard deviation in the typical study. In the absence of definitive research on IVD effectiveness, it is reasonable to expect improvements on this order of magnitude. A methodological implication is that research comparing IVD with an existing form of training should be designed to have sufficient statistical power to detect a "modest" difference, should differences be found.

At the same time, however, there is some reason to question the applicability of the meta-analyses to training outcomes, primarily because the outcome measures in the meta-analyses are largely measures of performance on written examinations. The criterion for evaluating the effectiveness of military training is commonly accepted as job proficiency, which is customarily measured in tests of hands-on performance.[53] Such measures must be included in any credible evaluation of IVD training effectiveness. However, we cannot be absolutely certain of the size of the improvement to expect on such measures.

Effects of IVD would also clearly depend on the type of IVD application. Where IVD supplements existing hands-on training, we expect that the effects should be positive, with a minimum expectation for a "modest" improvement in proficiency. This expectation would not necessarily apply to situations in which IVD replaces hands-on training, however. Proficiency could possibly improve, but it could also possibly decrease if IVD were substituted for some hypothetically necessary amount of hands-on training. Indeed, a wholesale substitution of IVD for hands-on training, as was done in some earlier studies, may be undesirable, except where there may be no hands-on opportunity.[54] Rather, a possibly beneficial application would partially substitute IVD for hands-on training, within a mix of training resources. Such a mix of resources could be less costly than

---

[53]A. K. Wigdor and B. F. Green, *Assessing the Performance of Enlisted Personnel: Evaluation of a Joint-Service Research Project*, National Research Council, Washington, D.C., 1986.

[54]Such a situation would be impossible to evaluate, given the lack of a comparison condition.

current training consisting exclusively of hands-on training, while providing equally proficient trainees.

In addition to task proficiency, it is desirable for an evaluation of IVD effectiveness to account for training time. As seen in previous evaluations of IVD technology, the amount of practice received on IVD is necessarily confounded with the method of training (supplementation or substitution). If IVD is expected to increase the efficiency of training, then a full evaluation should measure either the increase in practice that is afforded within an existing block of instruction, or, alternatively, it should measure the time required for individual trainees to achieve competency at the criterion. The Army is not now oriented toward individualized or self-paced instruction in its advanced individual training courses, where IVD is most commonly used. Thus, the amount of training is measured within existing blocks of instruction, to account for the effects of IVD and hands-on practice.

Summary

Our review of research has led us to conclude: (1) experimentation is the most appropriate method for providing precise estimates of the training effectiveness of interactive videodisc technology, (2) the principal training applications that should be examined are *supplementation* of hands-on training with IVD and *substitution* of IVD for expensive hands-on training, and (3) the appropriate criteria for experimental analyses are training time and job proficiency.

After reviewing the various advanced individual training courses using IVD at the Signal Center, we found two courses that each provided an opportunity to test one of two forms of IVD training: MOS 31M (Multichannel Communications Equipment Operator), for testing the use of IVD as a supplement to hands-on training; and MOS 31Q (Tactical Satellite/Microwave Systems Operator), for testing the use of IVD as a substitute for hands-on training. The next sections of this report describe the research design that was employed in each of these two courses, as well as the results that emerged.

III. SUPPLEMENTATION EXPERIMENT: THE 31M COURSE

OVERVIEW

The first phase of this research was a controlled experiment in the Signal Center's initial training course for Military Occupational Specialty 31M, Multichannel Communications Equipment Operator. There were several reasons for particular interest in the 31M application of interactive videodisc. First, 31M is one of the largest occupational specialties in the Signal Corps, accounting for more than 7500 members of the active Army. At the time this experiment began (1986), Fort Gordon was training about 2000 personnel per year as new 31Ms.

Second, the 31M occupational specialty was the first MOS for which the Signal Center developed its own IVD courseware. Signal Center managers created this courseware because they felt a need to impart equipment-specific information about a wide variety of equipment in a short time.[1] Both field units and instructors in the course at Fort Gordon had expressed this need. As a result, a number of interactive courseware products had been developed for the 31M specialty, and their use was well-accepted and institutionalized in several parts of the school curriculum.

Third, the Signal Center already tended to use IVD to *supplement* other instruction, an application that is typical of most Army uses of IVD. In the 31M course, as in many courses where equipment is expensive and in short supply, each classroom had many more students than pieces of communications equipment on which trainees could practice. During periods of "practical exercise"—an important and time-consuming portion of most Army courses—students were expected to use actual components and assemblages to learn how to set up, operate, and troubleshoot equipment. In effect, students had to wait for their turn on the equipment, creating "dead time" when IVD could be used for extra practice.

---

[1]31M personnel may be assigned to several types of units whose communication gear varies. The Advanced Individual Training course at Fort Gordon has to cover several different types of equipment, and the instructors believed that time allotted was insufficient for thorough training of some tasks. Thus an early and important objective of the IVD courseware, in the Signal Center's eyes, was to improve the *efficiency* of training, or the extent to which student time was appropriately used.

To examine the effectiveness of this type of IVD application in a rigorous and systematic way, we designed a controlled experiment in which 31M students were assigned to one of two equivalent groups: one group received training only on actual tactical equipment ("hands-on" training), whereas the other group received training both on tactical equipment and on IVD systems. The experiment lasted from June 1986 through January 1987. During each week, as students entered the course, RAND researchers assigned them to the two groups based on a randomization and balancing model. In all, 428 students participated, and we monitored the extent of training for each student (both hands-on and IVD training). Finally, near the end of the course we arranged for each student to be tested in a high-fidelity simulator that duplicated the face plates and actions of the tactical equipment and that provided standardized, computer-monitored performance assessments for every study participant. We next describe the characteristics of the 31M course and its use of IVD courseware, the experimental design, the implementation of the experiment, and the results.

DESCRIPTION OF THE 31M COURSE

Soldiers usually enter the 31M Advanced Individual Training (AIT) course immediately after basic training. Basic training normally lasts eight weeks; the 31M course requires an additional 14 weeks. Two-thirds to three-fourths of the students are new entrants to the active duty Army, and almost all of the remainder are new entrants to Army reserve components. Both groups are receiving MOS training for the first time; the great majority are young men with little or no previous exposure to the military or to communications equipment.[2]

The primary purposes of the course are to familiarize new soldiers with Army communications doctrine and equipment and to train them to perform most of the important tasks that they will need to know when they become part of an Army field unit.[3] The 31M job is complicated by the number and variety of types of equipment in the field. At a minimum, 31Ms must be able to handle three primary cat-

[2] A small fraction are prior-service personnel who have been in the Army before, or personnel who are retraining for the 31M MOS. They normally have a pay grade of E-4 or higher.

[3] Technically, the AIT course trains only a subset of "critical tasks" required of an MOS holder. Some tasks, particularly those that vary across units, are only partially trained in school or are trained entirely in the unit.

egories of equipment: electric power generators, antennas, and "communications equipment" (meaning electronic gear such as radios, multiplexers, telephones, their associated cables, and so forth). The communications equipment may be used in different ways, for example as a radio or cable system. The 31M must be prepared to operate a signal center either as a terminal point or as a relay point (between terminals); these functions involve different types of tasks. Finally, 31Ms will use one of three kinds of electronic equipment (low-, medium-, and high-capacity assemblages) and the different types of antennas and generators that accompany them.

Given this range of equipment in the field, the Signal Center teaches the 31M course in several segments. At the time we instituted the experiment, the major segments of the course included the following:

- Introductory material common to all communications equipment (two weeks)
- Medium-capacity equipment (five weeks)
- Low-capacity equipment (four weeks)
- Field training exercise (an outdoor exercise involving all aspects of the job, one week)
- Training in the Reactive Electronic Equipment Simulator (one week)
- High-capacity equipment and end-of-course comprehensive test (one week).

We established the experiment in the low-capacity equipment segment, during the eighth week of the course, when students were first introduced to the low-capacity assemblage, the AN/TRC-145.[4] The primary reasons for selecting this segment were that IVD courseware had been developed for the AN/TRC-145, the instructors were ready and able to use it, and the necessary IVD equipment was available for the two 31M classrooms in which low-capacity equipment was taught. Moreover, we found that the course's Reactive Electronic Equipment Simulator (REES), which was used in the thirteenth week of the course, could serve as a ready-made performance-testing device for low-capacity equipment (the REES is described in detail later in this section).

---

[4]The AN/TRC-145 includes a radio set (AN/GRC-103), two multiplexers (TD-660B/G and TD-754/G), a security device (TSEC/KG-27), and a telephone signal converter (CV-1548A/C), plus associated smaller items. For details see the technical manual, TM 11-5895-453-14.

Organization of Classrooms

The classes and classrooms were organized in a way that facilitated use of IVD as a supplementary training device. A new cohort of students entered the course each week, beginning in the week 1 classroom. When that cohort finished a week of instruction, it moved to a new classroom for the following week of instruction, and so on through the 14 weeks of the course. During week 8 of the 31M course, students received lectures and practice on the initial procedures for setting up the AN/TRC-145. Primarily these procedures consist of connecting proper cables, presetting switches and controls, and checking voltages.[5] Each classroom normally had about 25 students, who were taught by one senior civilian instructor and monitored by two military instructors. Before the advent of IVD, each classroom contained the instructors' area, a set of tables serving as student desks, and a set of 10 equipment assemblages consisting of stacks of components on racks. A majority of classroom time was devoted to practical exercises. During that time 10 students would be assigned to assemblages while the others worked at their desks, studying manuals or reviewing text material. As a student finished working at an assemblage, the instructor would check his work and then send him back to a desk, while another student would move to the assemblage. During a normal week, each student would be expected to practice several different installations at different assemblages. Instructors kept records to ensure that all students were rotated to assemblages in an equitable fashion.

The experiment was set up by adding eight IVD machines to one of the two week 8 classrooms.[6] Figure 3.1 illustrates the resulting complement of machines in the two rooms. In the experimental classroom, students were able to rotate among IVD units and equipment assemblages, thus effectively increasing the training stations from 10 to 18. Within both classrooms students were rotated in random order, and we monitored their practical exercises through a record-keeping system using cards. Each time a student was sent to a training station (either an assemblage or an IVD unit), the student filled out a card indicating the start time and other data. When he

[5]Near the end of the week students proceeded to operate an AN/TRC-145 in various modes, such as loopback and communication between different assemblages. However, the experiment focused on cabling and presets.

[6]IVD machines were also used for the experimental group in the week 9 classroom, in which more advanced AN/TRC-145 operations were trained. Although the courseware for week 9 contained some material relevant to the advanced tasks, we could not assess performance of them in the REES and therefore our analyses do not attempt to evaluate training for those functions.

left the assemblage, the instructor determined the ending time and certain other information.[7] The RAND research team provided an on-site research assistant for the duration of the experiment to monitor operations; the assistant visited the classroom frequently and ensured that training sessions and times were being promptly and accurately recorded. As we will show in the detailed discussion of results, this monitoring helped to establish that the IVD was used extensively and thus that the experiment was in fact implemented.

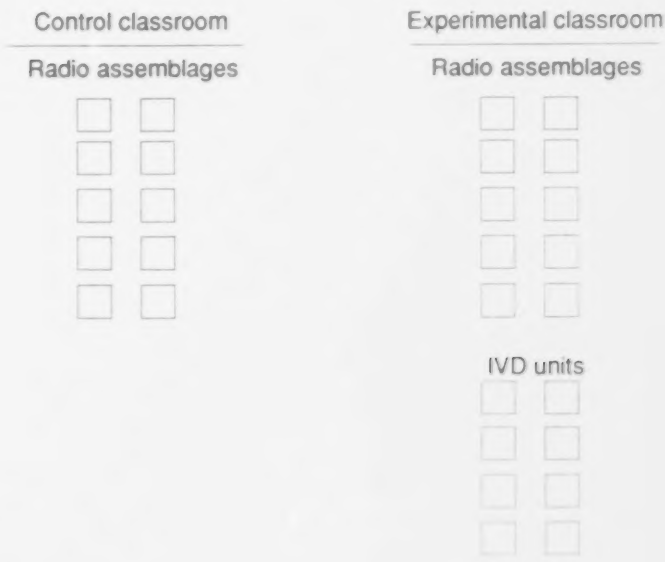

Fig. 3.1—Experimental and control classrooms

[7]For assemblage installation, the instructors attempted to record the presence or absence of an error in the installation, but as we observed classroom operations, we concluded that instructors were not sufficiently consistent in their assessments of the number and types of errors to warrant analysis of the error data.

EXPERIMENTAL DESIGN

Assignment of Students

A basic feature of a controlled experiment is that it enhances the likelihood that the groups being compared are equivalent. In many situations, equivalence can be ensured by assigning experimental units at random to the various conditions. Thus, we could have assigned each student to either the control classroom (without IVD) or the experimental classroom (with IVD) based on a coin toss or a table of random numbers. For large samples, this method minimizes differences among the groups on all types of preexisting variables, including *unmeasured* variables.[8]

In addition, it is often desirable to exercise direct control over the relative balance of the groups on specific variables that one knows (or believes) to be important. Intuitively, for example, if one believes that a student's initial level of electronics aptitude will affect his success in the course, one would like to ensure that the two groups are as closely balanced as possible on electronics aptitude. Furthermore, there is a statistical reason for preferring close balancing. With simple randomization, the sample statistics for relevant comparisons (such as contrasts in the mean performance levels between two groups) will be unbiased, but their variance will depend on the degree of balance among other variables that affect the outcome. If one ensures, in advance, that the groups are well balanced on such causal variables, the variance of a contrast will be reduced and the comparisons rendered more precise. We achieved this balance by using a method previously developed at RAND for assigning experimental units to conditions and for evaluating the degree of balance on specific variables.[9]

The variables considered by our balancing method are shown in Table 3.1, which includes the 428 students in the experiment.[10] As

---

[8]Thus a randomization procedure is preferable to simple matching on a few variables. If the sample is randomized across experimental conditions, one can be confident, within the limits of random error, that the groups do not differ by more than a specified amount on any variable. See, for example, Henry Scheffe, *The Analysis of Variance*, John Wiley and Sons, 1959; and B. J. Winer, *Statistical Principles in Experimental Design*, 2nd edition, McGraw-Hill, New York, 1971.

[9]For discussion, see S. James Press, "The MISER Criterion for Imbalance in the Analysis of Covariance," *Journal of Statistical Planning and Inference*, Vol. 17, 1987, pp. 375–388; and J. Michael Polich, James N. Dertouzos, and S. James Press, *The Enlistment Bonus Experiment*, The RAND Corporation, R-3353-FMP, 1986.

[10]The test data are derived from records of the Department of Defense Military Entrance Processing Command (MEPCOM), which maintains demographic data on

Table 3.1

CHARACTERISTICS OF SAMPLE SUBJECTS IN MOS 31M

Item	Value
Demographic and background characteristics	
Percent male	93.2
Race distribution, percent	
White	65.9
Black	30.1
Other	3.9
Age distribution, percent	
17–18	26.2
19	27.6
20	12.6
21-22	17.7
23 or older	15.9
Pay grade (rank) distribution, percent	
E-1	76.4
E-2	7.9
E-3	12.1
E-4 or higher	3.3
Educational and aptitude characteristics	
Previous education distribution, percent	
Some college or more	1.2
High school diploma	84.2
GED certificate	3.6
Less than high school diploma	10.9
AFQT score, mean	55.9
AFQT category distribution	
I-II (65–99 percentile)	33.2
IIIA (50–64 percentile)	25.9
IIIB (31–49 percentile)	35.8
IV (10–30 percentile)	5.1
Electronics composite aptitude score, mean	107.1
Electronics information score, mean	54.2
Number of cases	428

military applicants and administers written and physical tests to qualify them for enlistment. The tabulations and analyses below omit reserve component personnel and persons for whom baseline balancing data were not available in MEPCOM records (mostly reserve personnel), since they were not balanced and were not considered part of this experiment.

is evident, the sample is overwhelmingly male, about two-thirds white, young (median age is 19), and junior in service (more than three-fourths were privates at the initial entry pay grade, E-1). Most have recently graduated from high school, although in this sample about 14 percent did not possess a high school diploma; very few had a college education. The majority scored above the 50th percentile on the Armed Forces Qualification Test (AFQT), a composite measure of general ability. We also obtained two other scores: a composite measure of "electronics aptitude" used by the Signal Center to gauge a student's general ability to succeed in electronics training, and a score of specific electronics information.[11]

Each of the above variables was used in the balancing model to assign students during each week. In brief, the model first assigned students to conditions purely at random, creating a candidate assignment (called a "design"). If there were, say, 50 students entering the course during that week, the candidate design would place 25 in one group and 25 in the other. Then the model evaluated the adequacy of the design by examining, for each variable, the difference in means between the groups. Table 3.2 shows these means and standard deviations for the entire sample. Based on experience with this type of model, we had preestablished minimum degrees of matching that we deemed desirable (e.g., the mean AFQT score for the experimental group was permitted to be no more than one point different from the mean score for the control group). The minima were established by evaluating a set of designs according to a statistical criterion called MISER ("minimum inflation of standard error").[12] If the candidate design did not meet the balancing criterion for each variable, the design was rejected and the model generated a new random design for reevaluation.

This process guaranteed that each week the incoming cohort of students was closely balanced on all measurable variables that could plausibly affect outcomes of the experiment. As Table 3.2 shows, in the aggregate the matching was very close indeed. For example, the mean AFQT score was 56.1 in the control group and 55.8 in the experimental group, whereas the within-group standard deviation (S.D.) is approximately 18. In addition, the use of randomization permits us to rule out, as the sample size increases, the possibility

[11]The AFQT score and the two electronics scores are derived from the Armed Services Vocational Aptitude Battery, the written test used to screen military applicants.

[12]See Press, 1987.

Table 3.2

BALANCING VARIABLES, MOS 31M

Variable	Control Group		Experimental Group	
	Mean	S.D.	Mean	S.D.
Race (proportion white)	.659	.475	.659	.475
Sex (proportion male)	.924	.265	.940	.238
Education (proportion high school graduate)	.839	.369	.866	.341
Rank (proportion E-4 or higher)	.024	.152	.046	.210
Age (mean)	20.5	3.3	20.2	2.8
AFQT score (mean)	56.1	18.2	55.8	18.3
AFQT category (proportion I-IIIA)	.582	.494	.599	.491
Electronics composite aptitude score (mean)	107.3	9.9	106.8	9.9
Electronics information score (mean)	54.3	6.5	54.4	6.7
Number of cases	221		217	

that the experimental and control groups may have differed substantially on other variables that could not be measured. Thus this design allows one to make inferences that are unlikely to be affected by confounding with variables such as race, educational background, military experience, or aptitude.

Implementation

Experimental research, particularly in studies of the introduction of new educational technology, often raises issues of implementation. Typically, an experiment is designed to provide a clear contrast between a traditional method of instruction and an innovative method. However, researchers often observe that classroom instructors do not carry out the procedures as planned; if so, the experimental intervention may fail to affect crucial intervening variables (such as learning opportunity) that played key roles in the original design of the new approach. If such an implementation failure occurs, the experiment may not shed much light on the effectiveness or the potential of the experimental approach. In this case, the issues revolve largely around the extent to which the IVD devices and tactical equipment were actually used in the classrooms.

To some extent, this study's history and design features worked to build in faithful adherence to the experimental purpose. The section of instructors teaching low-capacity equipment in the 31M course were already supporters of IVD; they had initiated the request for interactive courseware in the first place, they had advised the courseware developers on the subject matter, and they had worked closely with the technicians at Fort Gordon in designing specific pictures, interactive branching patterns, and other features of the IVD material. In addition, the study's research assistant was present much of the time, attending to details, assisting in collecting data, and ensuring that events proceeded as planned. This provided a qualitative "implementation check."[13]

Evidence of Implementation

To obtain quantitative evidence of implementation, we tabulated data from the in-classroom records that represented each student's practical exercises on IVD and tactical equipment. Table 3.3 shows the results, in terms of the number of training sessions and the number of minutes devoted to them. It distinguishes "hands-on" assemblage training sessions from IVD sessions (in the control classroom, of course, the IVD data are zero by definition). Three points can be concluded from these results. First, the experiment was indeed implemented, in that IVD was extensively used in the experimental classroom. Students in the experimental room used the IVD machines for an average of 7.4 sessions per student during week 8 of the course. The sessions accounted for 80 minutes of practical exercise time per student, or slightly more than 10 minutes per session.[14]

---

[13]A more difficult and subjective issue is the "quality" of the IVD courseware itself. Controversy abounds in the educational and trade press about the virtues of various approaches to IVD displays, interaction patterns, presentation methods, and other courseware features. We could not obtain any definitive measures of courseware quality, but we have informally reviewed much of the existing Army interactive courseware inventory, and in our judgment the 31M courseware is typical of most Army applications. Moreover, this courseware, like others in the Army inventory, was developed in accordance with Army standards, which include review to ensure compliance with given technical standards and specifications. Although it might not rank as "state of the art" in the eyes of courseware design specialists, more advanced or complex applications are also likely to be much more costly and therefore are unlikely to be used by the Army on a wide scale.

[14]We did not attempt to tabulate the IVD computer records of student times and errors in detail, but spot checks of printouts from these records indicated that students were using the time in apparently productive ways, running through the cabling, preset, and installation procedures in the interactive courseware and responding to the courseware's prompts as needed.

Table 3.3

PRACTICAL EXERCISE TRAINING IN THE CLASSROOM

Training Activity	Control Group Mean	Control Group S.D.	Experimental Group Mean	Experimental Group S.D.
Number of exercise sessions				
Hands-on training	10.0	2.6	8.6	3.0
Interactive videodisc training	.0	.0	7.4	2.3
Total exercise minutes				
Hands-on training	150.3	42.8	137.8	44.6
Interactive videodisc training	.0	.0	80.0	26.1
Number of cases	209		217	

Second, use of IVD substantially increased the amount of practical exercise time available to students. In the experimental room, the average student received almost as many sessions on IVD (7.4 sessions) as on the assemblages (8.6 sessions). Thus, under the experimental condition the typical student had 16 opportunities to go through system installation procedures, either on IVD or equipment, as compared with 10 opportunities under the control condition. In terms of total training time, the extra opportunity allowed by the addition of IVD resulted in a 45 percent increase (217.8 total minutes, counting hands-on and IVD training in the experimental group, as compared with 150.3 minutes in the control group). Thus, students in the experimental group received considerably more opportunity to practice procedures.

A third implication is that student time set aside for practical exercise was used more efficiently in the experimental group. Both groups spent the same amount of time in the school (about 40 hours per week). The experimental classroom, however, permitted a 45 percent increase in actual practical exercise time within a constraint of constant student time. Army training philosophy places a high value on such practical exercises, which are generally viewed as much more productive than time spent studying or receiving lectures. If that assumption is valid, then the addition of IVD presumably allowed the class to make more efficient use of available student time. This greater efficiency by itself, of course, does not demonstrate that the increased time ultimately improved proficiency (a subject to be exam-

ined below), but it does suggest that the IVD application was well received and extensively used in the training environment.

Close inspection of Table 3.3 reveals that the experimental intervention was largely, but not entirely, an add-on (supplementation) phenomenon. In the experimental group, where IVD was available, students trained somewhat less on the tactical equipment (8.6 vs. 10.0 sessions). This difference probably was due to the difficulties in rotating students among a greater number of stations and monitoring their work at them. It resulted in a small reduction in hands-on training time for the experimental group (about 8 percent, or 137.8 minutes vs. 150.3 minutes). Thus, the experiment entailed a small amount of substitution of IVD for hands-on training time. In the analysis below, we will use each student's data on hands-on training time to control for this phenomenon in order to estimate pure supplementation effects more accurately.

PERFORMANCE ASSESSMENT

The 31M course included a natural mechanism for automated performance assessment in the REES simulator. During the thirteenth week of the course—about four weeks after the experimental intervention—all 31M classes were scheduled for supplemental training in the REES. Under the normal Program of Instruction, students were able to practice system cabling and presets, installation, system operation, and troubleshooting during their week in the REES. We modified this procedure so that during the experiment the first two days of each cohort's REES time were devoted to a performance test of tasks trained in the IVD experiment.

The REES is a one-of-a-kind simulator containing four "nodes" or signal centers, each located in one corner of a large building at Fort Gordon. Each node contains seven communications assemblages, including the AN/TRC-145 and several other devices that in a tactical environment would be operated by members of other occupational specialties. The assemblages are stacked as they would be in a tactical shelter (which in the field would be moved on a truck), and their face plates contain switches, controls, and dials that duplicate precisely the appearance and function of real equipment. They also permit attachment of cables that would be used in the field. The signals, however, are transmitted to a central computer instead of to antennas or other communications devices. The computer records each switch action, evaluates student errors, and permits assemblages to communicate with each other in configurations that represent typical

field layouts. The console operator can display and monitor the status of each assemblage, as well as obtain summary records of a student's actions on specific tasks. For training troubleshooting, the computer permits the console operator to insert faults into the assemblages and to monitor the speed and accuracy with which students isolate the resulting problems.

We employed the REES as a testing device by stipulating specialized procedures for all cohorts of students in the experiment.[15] At the beginning of each cohort's week in the REES, students were given an introductory briefing on procedures and purposes of the test. They were shown the basic REES operational procedure for a given task: The student first logs onto the REES computer by entering an identifying number into a small console attached to the assemblage; he then proceeds to set up cables and manipulate switches to perform a given task; when he believes he has completed the task, he depresses a "task stop" button. At that point, if the computer detects an error in his set-up or switch actions, a red light is illuminated on the console. (The result is also displayed on a screen at the operator's console.) The student may then back up, reset switches or perform other operations, and again depress "task stop" to indicate another attempt at the task. Each depression of "task stop" was treated as a task trial; the fewer the number of trials required, the better the performance. In addition, the REES detects conditions that pose hazards to the equipment or the operator (such as improper cable connections that would damage the components) and indicates them by a lighted signal.

In the normal 31M course, each student is initially assigned to one assemblage in the REES and directed to perform a specific task or set of tasks for that assemblage. After all students have completed their tasks or time has run out, they rotate to other assemblages. We preserved this procedure because of its familiarity to instructors and its ability to occupy students' time during the REES period.[16] The console operator and the REES instructors (one located in each of the four nodes) were trained to observe student performance but not to

[15]The experiment was conducted for 19 weekly cohorts, although maintenance problems and scheduling difficulties prevented REES testing of several cohorts. Altogether, 11 cohorts were tested in the REES—428 students in all.

[16]As a result, some students received more "practice opportunity" in using the REES than others before starting the cabling/preset task. For example, a student who began his REES day on the AN-TRC/145 received no initial REES practice time, whereas one who started on another assemblage and then rotated to the AN/TRC-145 could receive, say, 80 minutes of experience with the REES (although with a totally different assemblage). As noted below, we monitored these times and adjusted for them in the analysis.

coach them during the test. The operator and instructors also recorded information on occasional hardware and software malfunctions. A RAND research assistant visited the facility daily during the test to monitor these procedures and to ensure the integrity of the testing mechanism.

RESULTS

We obtained data on these tests, including measures of performance, from the summary records maintained by the REES software. Of the 428 persons who were in cohorts tested in the REES, the summary records included valid test results for 79 percent of the sample, or 340 persons. The remaining "untested" persons include 36 students whose installation was affected by a computer malfunction, precluding a valid test; and 52 students who were not tested on this task in the REES. Some of the latter group were tested on other assemblages or tasks but were never rotated to the proper assemblage for the cabling/preset test, and some were absent from the REES during the days when their cohort was tested. We examined the background characteristics of the above groups, including the variables used for balancing, but found no statistically significant differences between the tested and untested groups. Thus we concluded that the tested group was not biased in any observable ways.

Our analyses focus on three primary performance measures of the 340 students who were tested on the REES cabling/preset task:

- Amount of time required by the student to complete the task
- Number of trials required to complete the task[17]
- Presence of one or more procedural errors during the task.

Other possible performance measures in the REES proved to lack sufficient variability for analysis. For instance, of the 340 students tested, 328 (96 percent of the sample) eventually managed to complete the cabling/preset task correctly. Thus, it was not feasible to evaluate overall ability to perform the task without considering time or effort required. Similarly, the REES computer tracked the occurrence of hazardous conditions, but such conditions affected only two

[17]A trial was represented by a "task stop" record, indicating that the student believed he had correctly installed the system. Multiple trials in the field could necessitate time-consuming reiteration of installation procedures and could require a noncommissioned officer to travel to the site to effect the installation.

students. Probably as a result of low variability, neither of these measures was related to experimental condition or to background characteristics in any of the statistical models that we examined.[18]

To determine possible effects of the experimental IVD program on our three primary outcome measures, we carried out regression analyses predicting each outcome as a function of experimental condition and other variables that could not be fully controlled in the design. For each model we included indicator variables (scored as one or zero) or continuous variables representing the following:[19]

- Experimental condition (an indicator for experimental group vs. control group)
- Sex
- Race
- Age (in years)
- Education (possession of a high school diploma)
- Electronics aptitude (as scored on the Army's entrance test)
- Shift during which the test was conducted in the REES
- Amount of practice time in the REES on other assemblages before the tested task (in minutes)
- Number of hands-on training sessions during the experimental intervention in week 8 of the course.

The various models require somewhat different functional forms depending on the nature of the dependent variable. We used ordinary least-squares (OLS) regression for the measures of task completion (time to complete and trials to complete) because those measures are continuously distributed. For presence or absence of an error, which is a zero-one indicator, we used logistic regression.

Amount of Time and Number of Trials

Table 3.4 shows the results of the OLS regressions. It indicates that, controlling for all other factors in the model, the amount of time required to complete the task by experimental group students was

[18]One other possible measure—knowledge of radio installation procedures—was available from a multiple-choice test developed by the course instructors, and it did have substantial variability although its items were only generically related to the cabling/preset task. However, knowledge as measured by this instrument was unrelated to other student characteristics including experimental condition.

[19]Numerous other variables were examined in alternative models, including indicators for each cohort and all of the factors balanced in the design, but none of them had significant effects on the results.

Table 3.4

REGRESSION RESULTS FOR TASK COMPLETION MEASURES

Variable	OLS Regression Predicting Time to Complete[a]			OLS Regression Predicting Trials to Complete[b]		
	Coefficient	Standard Error	t	Coefficient	Standard Error	t
Experimental group indicator	−1.792	.769	2.33[c]	−.305	.131	2.34[c]
Sex (male)	1.112	1.424	.78	−.498	.243	2.05[c]
Race (white)	−.854	.859	1.00	−.086	.146	.59
Age	−.011	.126	.09	−.002	.021	.10
Education (high school graduate)	.458	1.093	.42	−.146	.186	.79
Electronics aptitude score	−.040	.040	.98	−.007	.007	.97
Shift of REES test (prime-shift vs. off-shift)	.644	.766	.84	−.063	.130	.48
Practice time on REES before REES test	−.029	.005	6.18[c]	−.001	.001	1.67
Number of hands-on training sessions in week 8	−.297	.138	2.15[c]	−.049	.023	2.10[c]
Intercept	25.309	5.368	4.72[c]	4.072	.913	4.46[c]

NOTE: Ordinary least-squares models, based on 328 cases (all students who completed a test in the REES).

[a]Model significant at $p < .001$ ($F = 5.629$).

[b]Model significant at $p < .05$ ($F = 2.136$).

[c]Parameter significant at $p < .05$.

significantly lower than the amount required by control students.[20] The model also suggests that significant effects can be attributed to the amount of pretest practice time in the REES and to the amount of hands-on training time each student received during week 8. The results are similar for predicting the number of trials required to complete the installation. Again, the experimental group performed significantly better than the control group (requiring fewer trials to complete).

[20]Means of groups (standard deviation in parentheses) for time to complete (in minutes) were as follows: control group, 17.5 (9.5); experimental group 16.2 (8.2). For number of trials to complete, means and standard deviations were: control group, 2.0 (1.3); experimental group 1.8 (1.0).

Presence of Procedural Errors

Table 3.5 shows coefficients from the logistic regression predicting the presence of any error during the installation process.[21] Although the effects are generally weaker, they operate in the same direction as those in the preceding models.[22] For this model the coefficient for the experimental group has a smaller t-value, which approaches but does not quite meet the conventional .05 probability level for a two-tailed test.[23]

Table 3.5

REGRESSION RESULTS FOR PRESENCE OF REES ERRORS

Variable	Coefficient	Standard Error	t
Experimental group indicator	−.430	.239	1.79[a]
Sex (male)	−.499	.477	1.04
Race (white)	−.406	.272	1.49
Age	.078	.045	1.73
Education (high school graduate)	−.172	.338	.51
Electronics aptitude score	−.009	.012	.74
Shift of REES test (prime-shift vs. off-shift)	−.343	.240	1.43
Practice time on REES before REES test	−.004	.001	2.49[b]
Number of hands-on training sessions in week 8	−.042	.043	.99
Intercept	1.832	1.709	1.07

NOTE: Logit model, based on 340 cases (all students with REES tests). Dependent variable is an indicator for the presence of an error during the REES installation task. Model is significant at $p < .05$ (Chi-square = 21.39).

[a]Parameter significant at $p < .07$ (two-tailed test).

[b]Parameter significant at $p < .05$.

[21]A logistic regression is a more appropriate functional form than OLS when the outcome measure is a dummy variable. It permits interpretation of the predicted values as the probability of achieving a success on the dummy variable. The equation is of the form $y = 1/[1 + EXP(-\mathbf{x})]$, where y is the outcome variable, EXP is the exponentiation function (base e), and x is a linear combination of independent variables and their coefficients to be estimated.

[22]The percentage of group members making one or more errors were as follows: control group, 65.9 (standard deviation of 47.6); experimental group, 57.2 (standard deviation of 49.6).

[23]It may be argued, however, that in this case a one-tailed test would be more appropriate, since we would not expect the addition of practice opportunity to reduce proficiency and we therefore should not test for it; if that argument were accepted, this result would surpass the conventional significance level.

Predicted Performance

For regression models of these types, particularly the logistic model that involves a nonlinear function, it is often easiest to interpret results when displayed as predictions from the model. Table 3.6 exhibits mean values predicted by the regression for the control and experimental groups. These estimates were obtained by evaluating each function at the mean for all variables except the experimental/control indicator (for which either a zero or one was substituted). The results represent the differences in performance that one would expect to observe based on the models (for the typical individual in the sample), provided that all factors except experimental condition are held constant.

For example, this analysis suggests that, on average, an experimental student would be expected to complete the task almost two minutes quicker than a control student (an 11 percent reduction in time). The other measures show similar performance improvements: experimental students would be expected to take fewer trials and to have a lower chance of making a procedural mistake (both effects representing a 15 percent relative change). The last measure is perhaps the easiest to interpret. It suggests that in a randomly selected group of students, if all other things were held constant, we should expect that 57 percent of students trained in the experimental group will make an error while installing an AN-TRC/145, compared with 67 percent in the control group.

These results are consistent and they accord with expectations. Whether they are substantively important is open to interpretation.

Table 3.6

PREDICTED VALUES FROM REGRESSION MODELS

	Mean Predicted Value from Regression Analysis	
Outcome Measure	Control Group	Experimental Group
Time required to complete installation (minutes)	16.3	14.5
Number of trials required to complete installation	2.07	1.76
Percent making one or more errors during installation	67.2	57.2

Many instructors and field personnel assert that a 15 percent reduction in errors or effort is difficult to achieve but it is meaningful in a unit environment. They also report that time is critically important to communications during a battle and therefore that faster installations may make a large contribution to battlefield success. There is, however, no systematic way to quantify a given proficiency improvement in terms of battle outcomes.

In the context of other research results on computer-assisted instruction, the training effects of this application of IVD fall in the low- to mid-range. Our estimated effects correspond to a proficiency change of about .25 standard deviations. That is certainly not trivial; it represents, for example, a change in the average student's score from the 50th to the 60th percentile over the control group's baseline (assuming a normal distribution). However, typical previous studies—although much less controlled than the present study—have estimated effects in the neighborhood of one-fourth to one-half of a standard deviation. One reason why the effects are so modest may be the amount of practice opportunity that was available to 31M students. Although the instructors perceived that opportunity was limited for some tasks, we observed that most students were able to practice AN/TRC-145 installation about eight times on actual equipment during the applicable training week. This hands-on training may have been sufficient to bring proficiency up to a point of substantially diminished returns.[24] It is possible that even in the control group, students were receiving so much hands-on training that the addition of IVD was unable to make a large difference in their performance. It may be true, therefore, that under conditions of more impoverished training opportunity the effects of adding on IVD would be greater.

[24]In fact, our inspection of instructor records of student errors during hands-on installation tasks during week 8 indicated that student error rates declined quickly during the first four to five installations and then leveled off thereafter.

IV. SUBSTITUTION EXPERIMENT: THE 31Q COURSE

OVERVIEW

The second phase of this research was a controlled experiment in the Signal Center's initial training course for Military Occupational Specialty 31Q, Tactical Satellite/Microwave Systems Operator. MOS 31Q is another of the pr imary specialties providing communications equipment operators in the Army; its members play vital roles in supporting battlefield command and control at the highest echelons of command. Although smaller in size than 31M, 31Q nonetheless contains a substantial number of personnel—approximately 1500 members of the active Army and an additional 1000 members of the reserve component belong to this MOS. When the experiment began in 1987, Fort Gordon was training some 750 personnel per year as new 31Qs.

As with the 31M MOS, the Signal Center developed its own IVD courseware for 31Q and extensively used the products. The reason for product development, however, is somewhat different from the case of MOS 31M. The Signal Center created the courseware because of an inability to train certain tasks and because expensive tactical equipment was in short supply. Only three radio assemblages were available for a class of 18-20 students, and several tasks could not be practiced because of possible danger to the equipment.

Consequently, the Signal Center obtained IVD hardware and developed four interactive videodiscs containing 17 tasks for use in the MOS 31Q course. The initial intent was to use the IVD courseware both as a substitute—to train students on the tasks they could not learn in any other fashion—and as a supplement—to train students on other tasks on the scarce equipment. Subsequently the course obtained additional radio assemblages, providing an opportunity to rigorously contrast training received on IVD with training received on actual equipment.

Thus, to examine the effectiveness of *substituting* IVD for more expensive training resources, we designed a second controlled experiment that capitalized on the availability of IVD and equipment. A key feature of this study is that it systematically compares the effects of using different mixes of resources to train two equivalent groups: one classroom employed only expensive radio assemblages, whereas the other employed a less expensive mix of tactical equipment and

IVD. The study's hypothesis was that the two groups of students would prove equally proficient. Using our randomization and balancing model, we assigned 31Q students to one of the two classrooms, monitored the amount and type of training received by the students, and assessed their performance using hands-on tests administered by objective assessors who were unaware of students' training experiences. The experiment lasted from September 1987 through July 1988. By the end of the study, 336 students had participated. We next describe the course, the experimental methodology, and the results in more detail.

DESCRIPTION OF THE 31Q COURSE

Soldiers train in MOS 31Q during AIT, subsequent to basic training. The entry standards for this occupation are the same as those for MOS 31M,[1] and the MOS also consists primarily of young men and women without prior military service. However, this course is longer and regarded as more difficult than the 31M course, and, as we shall see, the characteristics of the personnel trained in this specialty are somewhat different from those in MOS 31M.

During the 17 weeks of the course, the trainees learn their major responsibilities: installation, operation, and preventive maintenance of tactical satellite, microwave, and tropospheric scatter radios, multiplexing equipment, and their supporting antennas, generators, and communications security devices. Because 31Qs tend to be located at the signal centers at command posts of higher echelons (typically corps or army), their equipment tends to be more powerful and complex than that of the 31Ms. However, 31Qs are responsible for a smaller number of radio assemblages.

At the time of our experiment, the course was organized around the major categories of equipment as follows:

- Introductory material, including multiplexing equipment (three weeks)
- Tropospheric scatter and line-of-sight radios and related equipment (eight weeks)

[1]Entry standards for AIT courses specify minimum required scores on occupational subtests from the Armed Services Vocational Aptitude Battery (ASVAB), a test given prior to entry to military service. At the time of the study, both MOS 31M and MOS 31Q required that trainees possess a score of 95 or greater on the Electronics Composite aptitude scale.

- Tactical satellite radio equipment and related equipment (four weeks)
- Outdoor field training exercise (one week)
- End-of-course comprehensive testing and out-processing (one week).

We conducted this experiment during the fourth week of the course, at the beginning of the segment training the operation of tropospheric scatter (TROPO) and line-of-sight (LOS) radio equipment. The IVD courseware had been developed for this segment of the course; moreover, sufficient IVD players and tactical equipment were available for experimentation, and the IVD systems had been successfully implemented.

Development of IVD

The IVD courseware was developed for several reasons, but primarily to train tasks necessary to operate the radio terminal set AN/TRC-121,[2] of which a principal element is the tropospheric scatter radio set, AN/GRC-143.[3] At one time, there had been a shortage of TROPO radios for training in this section of the course, (three radios were available to teach a typical class of 20 students). The courseware was also developed because instructors were reluctant to train some tasks, principally so-called "alignments" and "adjustments" of the radio's constituent modules, fearful of damaging the equipment.[4] A third reason was that students were unable to practice power amplification because of insufficient power capacity in the classrooms.

The IVD courseware taught the use of the receiver and power amplifier alignments of the TROPO radio that could not be covered adequately in the class.[5] At the time we designed the experiment, this

[2]The AN/TRC-121 contains two radio sets (AN/GRC-143), a power supply (PP-4763A/GRC), two signal converters (CV-425/U), one radio set for use in alignment and while moving (AN/GRC-106), a telephone set (TA-312/PT), and two antenna groups (AN/TRA-37). For details, see technical manual TM-11-5820-602-15.

[3]The AN/GRC-143 is a general-purpose microwave FM radio set, operating over 12 or 24 channels, that uses tropospheric modes of propagation. It consists of a transmitter (T-961/GRC-143), receiver (R-1287/GRC-143), and radio frequency amplifier (AM-6090/GRC-143). For details, see technical manual TM-11-5820-595-12.

[4]The alignments and adjustments are performed whenever the equipment has been moved and as part of regular maintenance. They involve delicate adjustment, using tuning tools, of sensitive and expensive modules.

[5]The procedures were IF gain alignment, AGC alignment, squelch adjustment, APC alignment, receiver combined alarm adjustment, receiver combiner adjustments, power amplifier beam time delay, and power amplifier blower time delay.

block of instruction occurred in two parallel classrooms. In one classroom, students received lectures and considerable practice learning to operate the TROPO radio and the closely related LOS radio;[6] in the adjacent classroom, they learned troubleshooting of both radios. Each of the classrooms was managed by two civilian instructors and one military instructor.

EXPERIMENTAL DESIGN

A few months before this study began, the 31Q course acquired five additional TROPO radios, providing a total of eight TROPO radios in the course segment. The resources now available allowed us to organize the classrooms in a way that tested the substitutability of IVD for TROPO radio equipment. Although it was still not possible to train the tasks relating to the power amplifier on actual equipment, it was now possible to provide training on the TROPO receiver maintenance alignments and to contrast it with IVD training.

Organization of Classrooms

The experiment was set up by reconfiguring the two classrooms so that one contained seven TROPO radios and eight LOS radios (the control condition), while the other contained just one of each type of radio, plus eight IVD stations (the experimental condition).[7] The configuration of equipment in the classrooms is shown in Fig. 4.1. The arrangement was designed to create a sharp contrast in the cost of the available training resources (see below), while providing students in both classrooms with roughly equivalent practice opportunity. A typical class of 18–20 students was divided into two groups of 9–10 students, and each group had training stations on which to practice radio alignments. Students in the experimental classroom

[6]The LOS radio (AN/GRC-144) is part of the radio repeater set AN/TRC-138; it is a general-purpose microwave radio set intended for use in a 48-channel multichannel communications system using direct, point-to-point communication. It consists of a transmitter (T-1054/GRC-145) and receiver (R-1467/GRC-144). The principal difference between it and the TROPO radio concerns the receiver and the power amplifier. The LOS radio has no power amplifier and uses a single receiver, whereas the TROPO radio has an associated power amplifier and uses a dual receiver. The LOS radio is regarded by instructors as very similar to the TROPO radio but easier to use. For details, see technical manual TM-11-5820-695-12.

[7]The experiment focuses on receiver maintenance alignments for the TROPO radio, although the course instructors felt that training transferred readily between the LOS and TROPO radios on similar tasks (e.g., IF gain alignments). Thus, to avoid potential confounding between IVD and LOS training, we maintained an equivalent disparity between the groups in available LOS equipment.

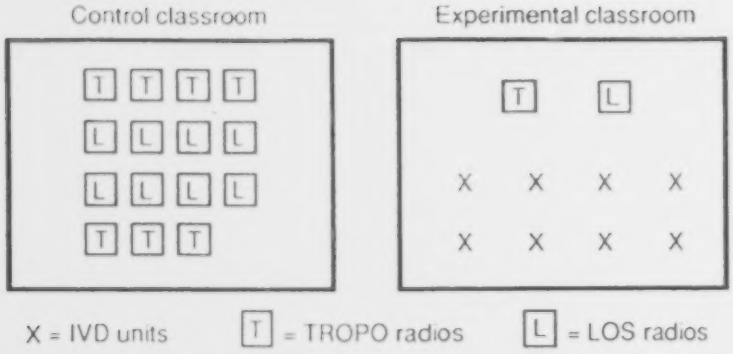

Fig. 4.1—Arrangement of equipment in TROPO/LOS classrooms

would practice on IVD and the two equipment assemblages (using the LOS radio for tasks in common), whereas students in the control classrooms would use only radio assemblages.

Costs of IVD and Equipment

According to documents filed by the Signal Center with the Army Training and Doctrine Command, the two training environments in this study represent significant differences in the acquisition and maintenance costs of the training resources in the respective classrooms. At the time the IVD was acquired, a new tropospheric scatter radio cost approximately $138,000. The costs of IVD hardware were reported at $5500 per system, and the Signal Center estimated that it cost approximately $40,000 to develop the IVD courseware in-house. Annual maintenance costs were estimated at approximately $1200 for each TROPO radio and about $500 for each IVD system.[8] If such figures are accurate, then the entire IVD training system in this study (eight hardware systems and the courseware) could be acquired for less than the cost of one TROPO radio. The substitution of eight

[8]Thus, if the simple acquisition costs of the training resources were considered, the TROPO radios used in the equipment room cost approximately $1 million, compared with a cost of about $250,000 for the IVD systems and the one TROPO radio in the experimental classrooms. If the LOS radios are also considered, the cost difference widens further. The costs of the LOS radios, however, were not included in the documentation submitted to TRADOC. Course personnel in MOS 31Q have estimated the cost of one LOS radio at approximately $24,000.

IVD systems for six TROPO radios, then, appears to represent a substantial savings.[9]

Sample Size Calculation

Determining an appropriate sample size is especially important in an experiment of this type, in which the research hypothesis is that groups receiving alternative forms of training would prove equally proficient. If statistical tests showed "no significant difference" on selected outcome measures, we might be tempted to conclude that the groups performed equally well. An alternative reason for a lack of difference, however, is that the statistical test lacked sufficient power to detect a true difference. In statistical terminology, such a problem is referred to as a "Type II error," or failure to detect a difference that in fact exists. Often, the reason for a lack of statistical power is the use of a sample size that is too small to permit a test statistic to surpass conventional thresholds of significance (e.g., alpha of .05).

Traditionally, one guards against the possibility of Type II errors during research planning by performing a "power analysis"[10] to determine the sample size that is needed to detect a given difference in group outcomes. We conducted a power analysis as part of our experimental design. The objective was to estimate the number of students needed in the experiment, while ensuring that our analysis would be likely to detect important differences in proficiency between groups of trainees. Thus, if none were found, we would be confident that the groups' performance was in fact equivalent.

Power analysis requires assumptions about the size of the group difference and the probability of detecting the difference at given levels of statistical significance. In this study, where IVD replaced the "preferred" technique of equipment training, we wished to enhance the chance that should IVD substitution diminish performance, this would be found. Thus in our power analysis, we considered a true decrement of .30 standard deviation on our outcome measures to be sufficiently important to detect. If a difference of that magnitude existed, we wished the probability to be no less than .90 that a standard comparison of the treatments (with a two-sided *t*-test and prob-

[9]The precise cost comparison would depend on certain key assumptions, including but not limited to the costs of courseware development and the appropriate life-cycle of IVD systems compared with equipment. Moreover, the "savings" should represent true and not hypothetical cost avoidance. For the purpose of this study, we assume that evidence of IVD substitutability would result in the release of equipment from the training classroom to field units or the replacement of equipment by IVD training systems in a programmed acquisition.

[10]See Cohen, 1977.

ability of .05) would reject the hypothesis of equivalence. The results of our analysis indicated that a sample size of 375, split between the treatment and control group, would be adequate given these assumptions. Our final sample consisted of 336 students, which produced slightly lower power than desired.

Assignment of Students

We assigned students to the training classrooms using the same balancing and randomization model used in the 31M experiment. The variables are shown in Table 4.1, which includes the 336 trainees in the experiment. As in 31M, the sample is largely male, young, and white, but compared with 31M, this population has relatively more members of the higher AFQT categories and high school graduates.[11]

Each of the variables in Table 4.1 was used in the balancing model to assign students to groups; the results of our assignments are shown in Table 4.2. As can be seen, the experimental and control groups are closely balanced on each of the variables of interest. In no case does the difference between the means or proportions of each group on any variable even remotely approach conventional levels of statistical significance, as expected.

Training Procedures

Tasks Examined. Our experiment covered all of the IVD programs available for training TROPO receiver maintenance alignments, but we focus particular attention on three: the IF gain alignment, AGC alignment, and squelch adjustment.[12] These tasks were selected, in consultation with course experts, as the more important of the tasks. They are also sufficiently complex to detect differences in student proficiency and can be assessed using existing course equipment in a reasonable amount of time.[13]

---

[11]In this experiment, we included members of the reserve component as part of the study design. Our primary hypothesis—that the groups receiving alternative training would prove equal in proficiency—requires that we maximize the number of individuals in the study design for the strongest possible test of equivalence. Because the training course throughput was modest (approximately 20 students entering the course every other week), and many of the students were reserve personnel, we included them in the analysis.

[12]These are three of the defined tasks for which 31Qs are responsible (tasks 113-591-5006, 113-591-5007, and 113-591-5008). See Department of the Army, *Soldiers Manual MOS 31Q*, STP 11-31Q1-SM, September 1987.

[13]Two of the tasks (IF gain alignment and AGC alignment) were chosen for examination because they were considered the most important receiver maintenance alignments, as well as the more difficult procedures to perform. The third task (squelch adjustment), also important to radio communications, followed the others in

Table 4.1

CHARACTERISTICS OF SAMPLE SUBJECTS IN MOS 31Q

Item	Value
Demographic and background characteristics	
Percent male	84.9
Race distribution, percent	
White	75.3
Black	20.5
Other	4.2
Age distribution, percent	
18	13.9
19	31.6
20	15.1
21–23	22.2
24 or older	17.2
Pay grade (rank) distribution, percent	
E-1	70.5
E-2	9.9
E-3	18.1
E-4 or higher	1.5
Military component, percent	
Active duty	81.2
Army reserve	6.9
Army National Guard	11.9
Educational and aptitude characteristics	
Previous education distribution, percent	
Some college or more	7.0
High school diploma	89.1
GED certificate	1.3
Less than high school diploma	2.6

sequence. The other tasks were eliminated from primary attention, principally because of difficulties in developing an adequate performance test. The two power amplifier procedures, for example, could not be tested for lack of a power source in the classrooms.

Table 4.1 (continued)

Item	Value
AFQT score, mean	62.2
AFQT category distribution	
I-II (65–99 percentile)	46.2
IIIA (50–64 percentile)	33.1
IIIB (31–49 percentile)	20.4
IV (10–30 percentile)	.3
Electronics composite aptitude score, mean	109.1
Electronics information score, mean	53.9
Number of cases	336

Table 4.2

BALANCING VARIABLES, MOS 31Q

	Control Group		Experimental Group	
Balancing Variable	Mean	S.D.	Mean	S.D.
Sex (proportion male)	.836	.371	.862	.345
Race (proportion white)	.752	.433	.754	.432
Rank (proportion E-4 or higher)	.012	.110	.018	.133
Component (proportion active duty)	.825	.381	.781	.415
Education (proportion high school graduate)	.964	.115	.959	.160
Age (mean)	21.2	3.4	21.0	3.6
AFQT score (mean)	62.2	16.2	62.4	16.1
AFQT category (proportion I-IIIA)	.798	.403	.789	.409
Electronics composite aptitude score (mean)	109.0	10.0	109.1	10.3
Electronics information score (mean)	54.0	8.4	53.9	8.1
Number of cases	167		169	

Student Rotation. The students were assigned to one of the two groups near the beginning of the 17-week course, but the groups were not actually formed until immediately after the introductory lecture of the TROPO/LOS annex. After the split, control students were assigned to the classroom that contained the radios only, and experimental students were assigned to the room containing the two radios and eight IVD stations. Each of the two groups then spent approximately two days receiving practical exercise on equipment and IVD. In the control group, the students received all of their practice on the radios; they stayed at a single radio for the entire day, except if a rotation was required to make sure every student received practice on the TROPO radio (as might happen if the number of students exceeded the number of TROPO radios).

In contrast, students in the experimental group began their practical exercise listening to a brief introduction to the use of the IVD machines and the IVD programs. Students were then assigned to one of the IVD systems, except for two students who were assigned to one of the radios in the room. The instructor rotated students between the radios and the IVD systems, with two stipulations: (1) that all students be given hands-on practice on the most important tasks (IF gain, AGC) before anyone had hands-on practice on less essential skills and (2) that all students be exposed to all alignment tasks in at least the IVD format. Because there were only two radios in the room, students had to rotate fairly regularly throughout the day.

Monitoring Implementation. A major goal of our study was to assess the degree to which IVD could substitute for equipment. To address this question, we monitored the amount of training received by students on IVD and equipment in the two classrooms. We were also concerned, once again, with ensuring that the substitution of IVD was adequately implemented in the experimental classrooms.

As in the 31M experiment, we used independent research assistants (one in each room) to oversee the training provided to each group. They also collected data on the amount and type of practice received, using a system of "training session cards" similar to that in our study of MOS 31M. For each student training session, we recorded the type of training received (tactical equipment or IVD), the specific tasks that were practiced, and the total training time received during the session.

Table 4.3 shows the number of training sessions received by students, on average, using the available resources in the alternative classrooms. The data make apparent the extensive implementation of IVD in the experimental classroom. For each of the principal tasks, students received considerable IVD training. For example, in practic-

Table 4.3

TRAINING RECEIVED IN EXPERIMENTAL AND CONTROL CLASSROOMS
(Number of training sessions)

	Control Group		Experimental Group		
Task/Method of Training	Mean	S.D.	Mean	S.D	*t*-statistic
IF gain alignment					
Equipment	6.38	2.19	2.55	1.01	
IVD			3.45	1.34	
Total	6.38	2.19	6.00	1.82	1.74
AGC alignment					
Equipment	6.05	2.35	2.23	.96	
IVD			2.85	1.32	
Total	6.05	2.35	5.08	1.80	4.27[a]
Squelch adjustment					
Equipment	2.95	1.29	.45	.64	
IVD			2.21	1.02	
Total	2.95	1.29	2.67	1.26	2.09[a]
Total training sessions					
(All procedures)	26.22	10.19	22.34	6.78	4.11[a]
Total practice time (minutes)					
Hands-on	369.46	74.78	93.79	41.65	
IVD			260.04	55.92	
Total	369.46	74.78	353.84	66.55	1.98[a]
Number of cases	167		169		

[a] *t*-test of means is significant at $p < .05$.

ing the IF gain alignment, students received an average of 3.45 sessions on IVD and 2.55 sessions on the radios. Thus, they accomplished 58 percent of their training sessions on IVD. For practicing AGC alignment, the IVD provided 56 percent of the students' training sessions, and, for squelch adjustment, most of the training was IVD-based (83 percent of training sessions).

The data in Table 4.3 also imply that students in the equipment-rich classroom received more training sessions on the various tasks. The number of total practice sessions (regardless of method of training) is significantly greater for two of the principal tasks (AGC alignment and squelch adjustment) and for all of the receiver maintenance alignment procedures taken together. The groups also differ in total training time. The group trained with IVD received practical exercise totaling about 354 minutes, whereas the group trained exclusively

with equipment received about 379 minutes of training (a difference of 25 minutes over two days).

PERFORMANCE ASSESSMENT

We could not use the Reactive Electronic Equipment Simulator in this part of the study because the REES did not contain the TROPO radio. Thus, we took an alternative approach to performance assessment by developing and implementing a hands-on test.

Hands-on Test Development

We used the technical manual for the AN/GRC-143 radio and the *Soldiers Manual MOS 31Q* as the basis for documenting objectively how students performed on the three principal tasks. The *Soldiers Manual* specifies the step-by-step procedures required to complete each task. It provides time standards for the completion of each task and a checklist of "Pass" or "Fail" decisions for each step of the task. Using these criteria, we developed observation forms with the following additional feature: on each step, we further noted whether the soldier "passed" or "failed" on a "first" or a "later" try. These data were recorded to provide a sensitive measure of student errors. Thus, we could distinguish errors even when a soldier completed a task properly within the time allotted for the test. Procedural errors, though corrected, nevertheless decrease the efficiency with which the task is accomplished.

Training of Test Administrators. The hands-on tests were administered by independent and objective assessors who were selected, hired, and trained by the RAND research team. Three assessors were employed over the course of the study. All were retired military personnel from communications operation and repair specialties. Each test administrator was given extensive training, first in the performance of the tasks to be tested, then in use of the hands-on test form. Interrater agreement checks were conducted using other members of the experimental team as test subjects, as part of training and periodically during the experiment. In eight agreement studies conducted during the course of the experiment, we found no cases where assessors disagreed whether a test subject had accomplished the task within the Army time standard. Interrater agreement on "pass" and "fail" judgments of the steps in the three tasks was also high (98 percent for the IF gain, 100 percent for the

alignment AGC, and 99 percent for the squelch adjustment), and consistent with comparable studies of rater agreement.[14]

Hands-On Test Administration. Throughout the experiment, test administrators were kept completely uninformed of the training condition of the students they tested. Each student was tested by a single administrator. In addition, students were counterbalanced by group as they were assigned to test administrators, thus ensuring that each test administrator tested approximately equal numbers of soldiers from the control and experimental classrooms.

Students received their performance test after the practical exercise period (generally, training was received on Thursday and Friday; testing occurred on Monday). On test day, students were first administered a written test (described below). Three students were then selected for hands-on testing in the radio room. At the completion of testing, each student returned to the classroom and another student was selected for testing. This process was managed by a RAND research assistant, who also ensured that students did not divulge information about the tests to waiting students.

Measures of Job Knowledge and Attitudes

In addition to the hands-on test, we developed and administered a questionnaire that contained measures of relevant job knowledge and attitudes toward training. Items measuring knowledge of the various TROPO radio alignment procedures, along with additional items concerning general knowledge of the TROPO radio, were developed by a subject-matter expert from the 31Q course. After pretesting, we selected 41 items for inclusion in the test. Additionally, we were interested in learning whether student attitudes differed toward the methods of training that they received. We developed a six-item scale of attitudes toward training in which item wording was varied so that three of the items were phrased negatively and three were phrased positively.

We do not regard these measures as primary indicators of the experiment's outcome. General knowledge of the radio is not as specific to the IVD intervention as the hands-on test, nor is it as policy-rele-

[14]L. A. Johnson, A. P. Jones, M. C. Butler, and D. Mann, *Assessing Interrater Agreement in Job Analysis Ratings*, Naval Health Research Center, Report No. 81-17, San Diego, California, 1981; M. H. Maier and C. M. Hiatt, *On the Content and Measurement Validity of Hands-On Job Performance Tests*, Center for Naval Analyses, CRM 85-79, Alexandria, Virginia, August 1985; and W. A. Nugent, G. J. Laabs, and R. C. Panell, *Performance Test Objectivity: A Comparison of Rater Accuracy and Reliability Using Three Observation Forms*, Navy Personnel Research and Development Center, NPRDC TR 82-30, San Diego, California, 1982.

vant. Attitudes are even less directly relevant. However, we wanted to assess a broad set of dimensions on which IVD or hands-on training might exhibit differences, and so we included these variables as ancillary measures.

RESULTS

Given certain unavoidable problems of scheduling (e.g., differences in time available to train each cohort because of holidays or other interruptions), we established priorities for hands-on testing: testing of the IF gain alignment was treated as most important, followed by the AGC alignment and the squelch adjustment. When short of time, a given cohort would not be tested on the squelch adjustment; occasionally, the AGC alignment would not be tested either. Thus, of the 336 students participating in the experiment, we conducted performance tests on 332 for the IF gain alignment, 305 for the AGC alignment, and 295 for the squelch adjustment. The two alternative training conditions were equally represented among the untested individuals. These students did not differ from the tested individuals on any of the demographic or educational variables used for balancing. Thus, the tested groups show no evidence of bias.

Within the tested groups, we analyzed two measures of performance on each of three tasks (IF gain alignment, AGC alignment, and squelch adjustment):

- Ability to complete the task successfully within the Army time standard; and
- The percentage of steps completed successfully on the first attempt.[15]

The first measure indicates whether the soldier could accomplish the task within the time allotted in the *Soldiers Manual*, while permitting the soldier to discover and correct any errors made during performance of the task. To receive a "pass" the student must eventually perform each step correctly. The second measure indicates the "efficiency" of task performance by accounting for success at the first attempt to perform each step in the task. In the field, initial errors would need to be identified and corrected before the alignment or adjustment could be completed. Presumably, as the initial steps are

[15]We used the percentage, rather than a count of errors, because the number of steps could differ depending on the equipment readings.

performed more accurately, less time and effort must then be spent to make the equipment operational.

The hands-on tests proved to be highly reliable for the purpose of group comparison. For each of the tests, we examined the intercorrelations among the scores (pass or fail) on each step on the first try and computed Cronbach's alpha, a measure of the internal consistency among items. For the IF gain performance test, alpha equaled .95. The AGC alignment performance test also proved reliable (alpha of .80), as did the squelch test (alpha of .88).

We again conducted regression analyses to examine the effects of receiving IVD "substitution" training, as opposed to only hands-on equipment training. For each of the three tasks, we analyzed the outcome measures using the appropriate functional form. We used logistic regression for predicting the likelihood of successful task completion within the Army time standard (a nominal measure, scored "1" if the student passed the task and "0" if he failed); we used ordinary least-squares multiple regression for predicting the percentage of steps accomplished correctly on the first attempt (a continuous measure). In each model, we predicted the outcome based on experimental condition. We also controlled for the following other relevant background and aptitude measures:[16]

- Experimental condition (an indicator for experimental group vs. control group)
- Sex (an indicator variable for male)
- Race (an indicator variable for white)
- Age (in years)
- Component (an indicator variable for active duty)
- Electronics aptitude score (from ASVAB scores at initial entry)
- Number of total training sessions on the specific task during the experimental intervention in the course (including all forms of equipment and IVD)
- Assessor (indicator variables for two of the three assessors).

Total training sessions were included in the models to provide a "purer" test of the substitutability of IVD for equipment. Unlike the 31M experiment, where we controlled for group differences in *hands-on* training opportunity to provide a more precise test of the effects of

[16]The model used in this experiment differs slightly from that used in the 31M experiment. We excluded education (possession of a high school diploma) because of insufficient dispersion (96 percent of the population were high school graduates). We also examined other variables balanced in the design as well as indicators for each study cohort, but found that none affected the results.

supplementary IVD practice, this model estimates the difference between training methods, holding total practice constant. As shown earlier, the control groups received more training sessions (except on the IF gain alignment), which might translate into increased proficiency.[17] The inclusion of total practice in the model, then, adjusts for possible effects of extra practice in the control classroom, and also estimates the effect of training condition, holding total training opportunity constant.

The indicators for test assessor were included to improve the precision of the models. Despite our regular monitoring, we saw evidence in the data that raters differed in their judgments. The differences introduced no bias because the students were counterbalanced by group, but they introduced measurement error. Therefore, we controlled for test administrator in the models to account for error variance attributable to the assessors.

We next give the results for student performance on the three tasks: IF gain alignment, AGC alignment, and squelch adjustment.

IF Gain Alignment

The IF gain alignment proved to be the most difficult task for trainees to accomplish; overall, only 34 percent of the examinees accomplished the task successfully within the defined standard (10 minutes). The passing rate was 34.8 percent in the control group (standard deviation of 47.8), compared with 33.3 percent in the experimental group (standard deviation of 47.3). The differences are not significant ($t = .27$).[18]

The task consists of 27 separate steps. The average success rate (first attempt) on each step was 75 percent. Although the control group performed somewhat more efficiently (i.e., succeeded at a higher percentage of the steps on the first try), the difference between the groups did not prove statistically significant.[19] Thus, the per-

[17]Alternatively, one might wish to estimate the effects of IVD substitution as accomplished within the overall training package and not control for training opportunity. We also examined models that did not include total practice on the task, but found little difference except for a lessening of the predictive power of the models. The effects of experimental condition were unchanged.

[18]In cases like this, it is important to know if the sample size is large enough to detect a meaningful difference. This is the issue of statistical power—the probability that a statistical test will reject the null hypothesis given that the two groups are in fact different. For the analyses in this section, we assumed that to be meaningful, a difference should be as large as at least .30 standard deviation. If that were the true difference, then the power of these analyses is approximately .80.

[19]Means of groups (standard deviation in parentheses) were as follows: control group, 77.1 (23.0); experimental group, 73.2 (31.0); $t = 1.32$.

formances of both groups of students on this test appear equivalent on both measures for this task.

The regression results of both measures of performance on the IF gain alignment are shown in Table 4.4. The models show that, after controlling for other factors, training method is not related to (1) the likelihood of completing the task to standard and (2) steps accomplished correctly during task execution.[20] The models also suggest that, as electronics aptitude increases, performance improves on both measures. In addition, fewer procedural errors occur as the number of total training sessions increases, and there are fewer procedural errors among the active duty population.

AGC Alignment

Students found the AGC alignment somewhat easier to accomplish within the prescribed time (15 minutes), although less than half the trainees could perform the task to Army standard (41 percent of the sample passed). The passing rate on the task favored the experimental group (43.5 percent; standard deviation of 49.7), compared with the control group (39.1 percent; standard deviation of 49.0), although, given the wide variance within groups, the difference is not statistically significant ($t = .78$).

The AGC alignment contains 18 steps. When we view the percentage of steps accomplished successfully on the first attempt, the control group performs significantly better.[21] On the first try, they correctly completed an average of 90.0 percent of the appropriate steps (standard deviation of 11.2), compared with the experimental group's average of 85.8 percent (standard deviation of 18.2). The t-value for this comparison is 2.43 ($p < .05$), indicating that the difference between means on this measure is statistically significant.

When other variables are controlled in the regression models (Table 4.5), training condition is unrelated to the probability of completing the task to standard, but it is related to procedural errors made during the performance of the test. The students trained in the IVD-intensive environment were more error-prone in their first attempt at the alignment (although they were still equally likely to suc-

[20]There is, however, a confidence interval associated with the estimate reflected in the standard error of the training condition coefficient. For the probability of task completion, the standard error implies that, with 95 percent confidence, we can conclude that the experimental and control groups differ by no more than 10 percentage points.

[21]The mean percentage correct across both groups was 87.9 percent.

Table 4.4

REGRESSION RESULTS FOR PERFORMANCE OF IF GAIN ALIGNMENT

Variable[a]	Ability to Complete Task to Standard[a]			Percentage of Steps Correct, First Try [b]		
	Coefficient	Standard Error	t	Coefficient	Standard Error	t
Experimental group indicator	−.014	.244	.00	−.034	.029	1.17
Sex (male)	.284	.383	.74	.018	.043	.42
Race (white)	.193	.300	.64	−.049	.035	1.40
Age	−.068	.061	1.11	−.005	.007	.64
Component (active duty)	.031	.312	.10	.092	.038	2.41[c]
Electronics aptitude score	.036	.013	2.83[c]	.004	.002	2.81[c]
Number of total training sessions on task	.097	.063	1.55	.023	.007	3.04[c]
Intercept	−4.377	1.760	2.49[c]	.103	.212	.49

NOTE: Values for assessor indicator variables are not shown.

[a]Logistic regression model, based on 326 cases with complete data. Model is significant at $p < .05$ (Chi-square = 19.67).

[b]Ordinary least 1=1 squares model, based on 326 cases with complete data. Model is significant at $p < .05$ ($F = 5.15$).

[c] Parameter is significant at $p < .05$.

Table 4.5

REGRESSION RESULTS FOR PERFORMANCE OF AGC ALIGNMENT

Variable[a]	Ability to Complete Task to Standard[a]			Percentage of Steps Correct, First Try [b]		
	Coefficient	Standard Error	t	Coefficient	Standard Error	t
Experimental group indicator	.098	.259	.37	−.047	.018	−2.67[c]
Sex (male)	.043	.387	.10	−.029	.026	1.15
Race (white)	−.184	.303	.61	.017	.021	.82
Age	−.157	.063	2.48[c]	−.002	.004	.40
Component (active duty)	.509	.334	1.52	.013	.023	.57
Electronics aptitude score	.046	.014	3.33[c]	.002	.001	2.41[c]
Number of total training sessions on task	−.103	.063	1.63	−.002	.004	.53
Intercept	−1.945	1.800	1.08	.680	.125	5.42[c]

NOTE: Values for assessor indicator variables are not shown.

[a]Logistic regression model, based on 299 cases with complete data. Model is significant at $p < .05$ (Chi-square = 36.99).

[b]Ordinary least squares model, based on 299 cases with complete data. Model is significant at $p < .05$ ($F = 2.28$).

[c]Parameter is significant at $p < .05$.

ceed eventually). For this task, as with the IF gain, electronics aptitude is a significant and consistent determinant of performance.

Squelch Adjustment

The squelch adjustment proved the easiest task; the overall passing rate in the two groups was 82.4 percent (within a 10-minute standard). The two groups' performance on this task was consistent with their performance on the AGC alignment—they proved equal in their ability to accomplish the task within the defined interval,[22] but the experimental group was significantly less "efficient" at accomplishing the 14 steps in the task.[23] These conclusions are supported by the raw means, as well as by the regression models, when other factors are controlled (Table 4.6). The *t*-value for experimental condition is quite small for the measure of task completion, but it is significantly negative for the measure of initial step accomplishment. Again, in these models improvements in both measures of performance are attributable to electronics aptitude, as measured by the ASVAB. Soldiers in active duty status also demonstrated improved performance in these models.

Job Knowledge and Attitudes

The questionnaire on job knowledge and attitudes was administered to 331 individuals before the hands-on test period. Because subscales of knowledge of specific procedures proved unreliable, we used the entire set of 41 items as a test of general knowledge of TROPO radios and maintenance alignment procedures. The scale that resulted was fairly reliable for the purpose of group comparison (alpha of .72). The reliability of the attitude measure also proved adequate for group comparisons (alpha of .60).

Performance on the knowledge test proved unrelated to the type of training received; both groups provided correct responses to some three-quarters of the items.[24] The attitude measure, however, showed the group trained in the IVD-intensive environment to express more negative sentiments toward the training they received. On a five-point scale, where higher values indicate more positive atti-

---

[22]Means of groups, by condition (standard deviation in parentheses) were as follows: control group, 83.6 (37.2); experimental group, 80.5 (39.7); $t = .67$.

[23]Means of groups, by condition (standard deviation in parentheses) were as follows: control group, 94.3 (13.5); experimental group, 89.2 (21.0); $t = 2.47$, $p < .05$ (a statistically significant difference).

[24]Means of groups, by condition (standard deviation in parentheses) were as follows: control group, 77.6 (8.9); experimental group, 76.9 (9.9); $t = .66$.

Table 4.6

REGRESSION RESULTS FOR PERFORMANCE OF SQUELCH ADJUSTMENT

Variable[a]	Ability to Complete Task to Standard[a]			Percentage of Steps Correct, First Try[b]		
	Coefficient	Standard Error	t	Coefficient	Standard Error	t
Experimental group indicator	−.203	.330	.62	−.049	.020	2.39[c]
Sex (male)	−.000	.467	.00	−.006	.030	.21
Race (white)	−.106	.392	.26	.020	.025	.80
Age	−.108	.079	1.37	.000	.005	.08
Component (active duty)	.766	.383	2.00[c]	.051	.026	1.99[c]
Electronics aptitude score	.039	.019	2.02[c]	.003	.001	3.02[c]
Number of total training sessions on task	.131	.137	.95	.011	.008	1.33
Intercept	−1.948	2.409	.81	.423	.147	2.88[c]

NOTE: Values for assessor indicator variables are not shown.

[a]Logistic regression model, based on 289 cases. Model is significant at $p < .05$ (Chi-square = 26.82).

[b]Ordinary least-squares model, based on 290 cases. Model is significant at $p < .05$ ($F = 4.33$).

[c]Parameter is significant at $p < .05$.

tudes, the average response in the control group was 4.04 (standard deviation of .60), and the average response in the experimental group was 3.80 (standard deviation of .61). Although both responses are to the positive side of the scale midpoint, they still differ significantly ($t = 3.62$, $p < .01$).[25]

Predicted Performance

For an experiment of this type, in which the study hypothesis posits that groups receiving alternative forms of training would be equivalent on an outcome measure, predicted performance is particularly important. The predicted values, and related estimates of effect

[25]We also examined the responses using our regression model, minus the indicator variables for assessor and using the measure of total training sessions on all tasks. The models revealed a negligible effect of training condition on overall knowledge, but it proved significant on the attitude measure, indicating that the members of the experimental group were less satisfied with the training they received ($t = 3.05$, $p < .05$).

size, help provide meaning for the group differences that were manifested.

Table 4.7 shows the predictions from the models presented in this section when each function is evaluated at the mean for all variables except the experimental/control indicator. The table suggests that there are no major differences between the groups on any of these measures. For example, the table shows absolute parity between the groups in the likelihood of completing the IF gain alignment. Although the model predicts that each group has a slight advantage in completing one of the two other tasks, the differences do not remotely approach conventional levels of statistical significance.

The measures of procedural errors, although favoring the control group, still suggest differences in predicted performance that are quite modest. To illustrate, this analysis suggests that in attempting an AGC alignment, control students would be expected to complete 90.4 percent of the steps in their first attempt, whereas the corresponding expectation for the experimental students would be 85.6 percent—a difference of about .9 step of the 18 steps in the task.

Table 4.7

PREDICTED VALUES FROM 31Q REGRESSION MODELS

Outcome Measure	Mean Predicted Value from Regression Analysis	
	Control Group	Experimental Group
Percent completing IF gain within Army time standard	33.9	33.6
Percent completing AFC within Army time standard	40.0	42.3
Percent completing squelch within Army time standard	83.8	80.8
Percent of steps correct first time for IF gain	76.7	73.3
Percent of steps correct first time for AGC	90.4	85.6
Percent of steps correct first time for squelch	94.1	89.3

In the now-familiar context of effect sizes, our estimated effects in this application of IVD range from –.08 to +.04 of a standard deviation on the three measures of task completion; they range from –.12 to –.31 of a standard deviation in measures of procedural errors. The former effects are negligible, and although the latter effects are negative, they are modest. The reasons for these differences are open to interpretation, which we will provide in the next section.

V. CONCLUSIONS

This report has presented the results of research conducted to evaluate the effectiveness of an interactive microcomputer/laser videodisc (IVD) training system used to facilitate training in a variety of military occupational specialties in the Army and, increasingly, in the other services. We have reported our results for two experimental studies. In both studies, we examined the training effectiveness of an IVD system used in advanced individual training of communications-electronics specialists at the U.S. Army Signal Center, Fort Gordon, Georgia. The research objectives have been, in general, to apply principles of controlled experimentation to evaluate the benefits of IVD technology and to define general conditions for its effective use in training. In this section, we present our conclusions from the research, covering, first, the benefits of the methodology and the specific IVD applications examined, and, second, the implications of our findings for wider application of IVD technology to military training.

BENEFITS OF EXPERIMENTATION

In both studies, we applied elements of classic experimental design to establish causal relationships between the method of training and the resulting performance. We examine five principal elements of the research method.

First, we defined alternative conditions of training (one for each study) that represent feasible policy options for the use of training resources. The options contrasted the status quo (hands-on equipment training) with potential applications of IVD technology (supplementation or substitution). Second, we implemented these options within existing courses, assigning students in a randomized, balanced fashion to one training condition or the other. Randomization is the critical element allowing unambiguous causal inference; along with adequate statistical power, it is a key factor for enhancing confidence in a finding of "no difference" in the 31Q experiment.[1] Third, we carefully monitored the training that was administered, both to ensure that the experimental intervention was implemented and to measure the practice opportunity important for

[1]W. H. Yeaton and L. Sechrest, "Assessing Factors Influencing Acceptance of No-Difference Research," *Evaluation Review*, Vol. 11, No. 1, February 1987, pp. 131–142.

interpreting the effects of training. Fourth, we gathered data using a sample size that was sufficient, in our estimation, to provide a fair statistical test of the effects of the training methods. Last, to compare the effects of the alternative methods of training, we rigorously assessed trainee performance on job-related criteria, subsequent to the training.

We have concluded from our experience that randomized experiments, conducted within existing military training courses, are feasible and useful methods for establishing the effectiveness of innovative training technologies. However, experiments can be costly and time-consuming; they require close monitoring and supervision; and they imply logistical changes for existing military routines.[2] Nonetheless, comparative experiments are the most defensible, rigorous, scientific method for providing quantitative estimates of the differences in outcomes arising from alternative policies and practices. When such information is required, as, for example, to justify a major expenditure of public funds for a resource of uncertain benefit, the practical barriers to experimentation can and should be overcome.

BENEFITS OF IVD FOR SUPPLEMENTATION AND SUBSTITUTION

The military services, and particularly the Army, have shown special interest in applying IVD technology for training purposes. The Army Signal Center and other Army schools adopted IVD for classroom training early, and they have heavily emphasized its use in teaching occupation-specific skills and procedures. This trend seems likely to continue; the Army, through its EIDS system, expects to make substantial investments in IVD-based training technology. The other services are expected to follow a similar path. All these actions are based on the belief that IVD or similar computer-based training devices or simulators can provide beneficial instruction and training.

Our experimental studies were designed to assess the potential benefits in a rigorous, quantitative way. We selected the two most common (and we think most important) types of application for analysis: supplementation, in which IVD is added to an existing baseline program of "hands-on" training; and substitution, in which IVD re-

[2]For example, in the 31M training course, the standard practice was to assign students to parallel classrooms based on where they were billeted. Thus, to implement randomization without sensitizing students, we defined the quarters to which students should be assigned. This required enlisting the cooperation of the responsible organization, which was outside the 31M course's chain of command.

places more expensive hands-on training resources used in the baseline program. Other applications of IVD are possible, of course,[3] and proper care should be taken in generalizing from the specific MOS training courses, tasks, and courseware examined in this research. However, the vast majority of IVD applications in the Army are straightforward cases of supplementation or substitution in training. The value of most of the IVD inventory will depend on how well and how efficiently such applications pay off.

The two experiments showed that IVD can be beneficial in the applications we examined. The supplementation experiment in the 31M course found that IVD, combined with traditional hands-on training, provided 45 percent more opportunity for students to practice their skills and reduced (by about 15 percent) the amount of time and effort they required to perform radio installations. It also reduced the chances that they would make an error during the process.

Similarly, the substitution experiment in the 31Q course found that IVD could provide student proficiency equivalent to that provided by expensive tactical equipment, as part of a less equipment-intensive training resource mix. We found strong evidence that students trained primarily on IVD were able to align their communications systems as successfully as those trained exclusively on actual equipment. In both cases, we have high confidence in these findings because the studies employed large sample sizes, randomized and balanced assignment of students to groups, and systematic, objective performance assessments.

Nevertheless, IVD did not prove to be a panacea or an unqualified success in these applications. In the 31M course, overall task completion was not affected by the addition of IVD, probably because of a "ceiling effect"—when unconstrained by time, 96 percent of the students were eventually able to install their systems. In the 31Q course, the substitution of IVD for actual equipment reduced procedural performance (the number of steps during an alignment which the trainee performed correctly on the first attempt). IVD-trained students were also less satisfied with their training experience, compared with the students who trained entirely on real equipment. These results may indicate that the experimental (IVD) students may not have been as efficient or as confident about their ability as their control counter-

[3]For example, proponents often cite the value of using simulation to train tasks that are too dangerous or impractical to train by hands-on methods (such as combat surgery or flying aircraft). Other potentially useful applications—the use of IVD for providing sustainment training or aiding job performance in units—were not covered in the research.

parts.[4] We note, however, that the 31Q comparison was a rather "extreme" test: the IVD classroom had only two radios for hands-on training, and the total cost of the IVD room's equipment was on the order of one-fourth of its conventional counterpart. We suspect that a less extreme contrast in resource mixes would have placed the IVD approach in an even better light. However, considerations such as these do caution against wholesale substitution of simulation in place of hands-on training.

IMPLICATIONS FOR IVD DEVELOPMENT AND APPLICATION

These studies raise important issues regarding how IVD technology may be used most appropriately in other training applications. Both studies suggest training conditions that may be important for enhancing (or minimizing) IVD effectiveness in situations of supplementation or substitution. They also suggest potential criteria for selecting courses and tasks for which IVD courseware may be developed.

Implications of the 31M Experiment

In describing the results of the experiment in MOS 31M, we observed that trainees were able to accomplish the task in the REES (on which they were tested) with little difficulty, and they appeared to receive ample hands-on practice during the relevant portion of the course. We speculate that practice at the task examined may have been adequate to achieve proficiency, despite the apparent shortage of equipment, possibly because the task was not overly difficult. We believe that the above factors are interrelated. If so, then the effectiveness of IVD supplementation may depend largely on the difficulty of the task, the amount of hands-on practice opportunity, and existing levels of proficiency.

The density ratio of equipment to students, a common basis for justifying the addition of IVD resources, does not by itself imply insufficient practice or inadequate proficiency. Rather, equipment density is meaningful only in relation to the number of students that are trained and the amount of time available to train them. These factors together determine the amount of practice that is received, which when combined with task difficulty determines subsequent profi-

[4]It is also possible that experimental students' satisfaction with training was diminished by awareness of their counterparts' training environment.

ciency. Therefore, we believe that all these factors need to be considered to indicate a need for additional training resources.

The addition of training resources is also frequently justified by the improvement it promises for classroom efficiency. Results of the 31M experiment indeed support the hypothesis that more time can be spent on "practical exercise" within an established block of instruction when IVD resources are added. Although instructors may prefer to keep students occupied in practical exercise, this activity does not *necessarily* offer an advantage over other forms of training. When training time is held constant, it must be shown that improved classroom efficiency yields gains in proficiency.

An alternative approach for improving the efficiency of training is to add IVD resources and, through the extra practice IVD can permit, to shorten the length of a training course. To our knowledge, this approach has not been implemented in Army advanced individual training. It may represent, however, a fruitful application for IVD. Our research did not investigate this possible application, although it has been the subject of other research.[5]

Implications of the 31Q Experiment

The 31Q experiment supports the hypothesis that IVD can reduce costs by substituting for more expensive equipment, but the results also suggest important factors to consider in implementing this approach. If the substitution of IVD simulation for hands-on training increases procedural errors, then the practical significance of effects on procedures must be weighed against the cost savings achieved by substituting IVD technology. We suspect (although we have no data to support our hypothesis) that certain minimum levels of hands-on training may be required to ensure competency and self-confidence among trainees. If so, training managers may seek to optimize the trade-off between hands-on training and IVD simulation by balancing the cost savings against the risk that diminished procedural efficiency may imply for the ability to achieve wartime missions.

Cost Considerations

From a policymaking perspective, perhaps the most important comparisons between training alternatives involve not just effectiveness, but also cost—particularly in situations such as these where proficiency was not dramatically affected by either form of training. The two experiments provide very different lessons. In the 31Q com-

[5]Orlansky and String, 1979.

parison, we observed that the groups' ability to accomplish the tasks were essentially equivalent, but lower costs were achieved with the use of IVD. This application (substitution to achieve cost savings) could possibly find much broader use in the services. The 31M experience appears much more typical of military IVD applications; here we observed modest performance improvements, but at greater cost. Ultimately defense managers need to judge the importance of the performance increases in these cases. We would argue that while such cases may be justified in particular instances, the burden of proof should fall on the IVD proponent, who should show that the increased proficiency is really needed and that it is worth the cost. As a corollary, we would argue that applications involving substitution should be given priority, provided that reasonable evidence can be obtained to establish a presumption of equivalent outcomes.

REFERENCES

Army Science Board, Office of the Assistant Secretary of the Army, *Final Report of the 1985 Summer Study on Training and Training Technology—Applications for AirLand Battle and Future Concepts*, December 1985.

Balson, P. M., D. G. Ebner, J. V. Mahoney, H. L. Lipert, and D. T. Manning, "Videodisc Instructional Strategies: Simple May Be Superior to Complex," *Journal of Educational Technology Systems*, Vol. 14, No. 4, 1985–1986, pp. 273–281.

Balson, P. M., D. T. Manning, D. G. Ebner, and F. R. Brooks, "Instructor-Controlled Versus Student-Controlled Training in a Videodisc-Based Paramedical Program," *Journal of Educational Technology Systems*, Vol. 13, No. 2, 1984–1985, pp. 123–130.

Bangert-Drowns, Robert L., James A. Kulik, and Chen-Lin C. Kulik, "Effectiveness of Computer-Based Education in Secondary Schools," *Journal of Computer-Based Instruction*, Vol. 12, No. 3, 1985, pp. 59–68.

Becker, Henry J., *The Impact of Computer Use on Children's Learning: What Research Has Shown and What It Has Not*, Center for Research on Elementary and Middle Schools, Johns Hopkins University, n.d.

Best, D., M. Caldwell, and L. Harrelson, "Video Disc Training," *Army Trainer*, Vol. 2, No. 3, 1983, pp. 7–9.

Boldovici, John A., "VIGS Evaluation," Memorandum, U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Knox, Kentucky, June 1986.

Bracey, Gerald W., "Computers in Education: What the Research Shows," *Electronic Learning*, November-December 1982, pp. 51–55.

Branch, Charles E., B. R. Ledford, B. T. Robertson, and L. Robison, "The Validation of an Interactive Videodisc as an Alternative to Traditional Teaching Techniques: Auscultation of the Heart," *Educational Technology*, March 1987, pp. 16–22.

Campbell, D. T., and J. C. Stanley, *Experimental and Quasi-experimental Designs for Research*, Rand-McNally Publishing Co., Chicago, 1963.

Clark, J. W., "Videodisc Training at Fort Gordon: A Practical Application," *The Videodisc Monitor*, July 1985, p. 10.

Clark, J. W., "Videodisc Training Workshop: Fort Gordon, Georgia," *The Videodisc Monitor*, May 1985, p. 10.

Clark, R. E., "Confounding in Educational Computing Research," *Journal of Educational Computing Research*, Vol. 1, No. 2, 1985, pp. 137–148.

Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*, revised edition, Academic Press, New York, 1977.

Davis, Barbara Gross, *Nebraska Videodisc Science Laboratory Simulations* (Executive Summary), The Annenberg/CPB Project, University of Nebraska, March 1985.

Defense Science Board, Office of the Under Secretary of Defense for Research & Engineering, *Report of the Defense Science Board, 1982 Summer Study Panel on Training and Training Technology*, November 1982.

Department of the Army, Headquarters, U.S. Army Signal Center and Fort Gordon, *Basis of Issue Plan for Electronic Information Delivery System (EIDS)*, Fort Gordon, Georgia, 1987.

Department of the Army, Headquarters, U.S. Army Training Support Center, *EIDS Primer*, Fort Eustis, Virginia, n.d.

Department of the Army, Headquarters, United States Army Training and Doctrine Command, *Electronic Information Delivery Systems (EIDS) and Interactive Courseware (ICW) Implementation Plan*, Pamphlet 350-XX, Fort Monroe, Virginia, January 1988.

Department of the Army, Headquarters, United States Army Training Support Center, *Supplementary Interactive Courseware User's Guide (ICUG) Bulletin*, Fort Eustis, Virginia, August 1987.

Department of the Army, U.S. Army Training and Doctrine Command, TRADOC Systems Analysis Activity, White Sands Missile Range, *Training Effectiveness Analysis: A Process in Evolution*, TRASANA Pamphlet 350-4, White Sands, New Mexico, August 1985.

Department of the Army, *Soldiers Manual MOS 31Q*, STP 11-31Q1-SM, September 1987.

Ebner, D. G., D. T. Manning, F. R. Brooks, J. V. Mahoney, H. T. Lipert, and P. M. Balson, "Videodiscs Can Improve Instructional Efficiency," *Instructional Innovator*, Vol. 29, No. 6, 1984, pp. 26–28.

Edwards, J., S. Norton, S. Taylor, M. Weiss, and R. Dusseldorp, "How Effective Is CAI? A Review of the Research," *Educational Leadership*, November 1975, pp. 147–153.

Glass, G. V., McGaw, B., and M. L. Smith, *Meta-analysis in Social Research*, Sage Publications, Beverly Hills, California, 1981.

Gorman, P. F., "Educational Technology: Yesterday, Today and Tomorrow," *Military Review*, Vol. 66, No. 12, December 1986, pp. 4–11.

Hagler, P., and J. Knowlton, "Invalid Implicit Assumption in CBI Comparison Research," *Journal of Computer-Based Instruction*, Vol. 14, No. 3, September 1987, pp. 84–88.

Hasselbring, Ted, K. Fleenor, R. Sherwood, D. Griffith, J. Bransford, and L. Goin, "An Evaluation of a Level-One Instructional Videodisc Program," *Journal of Educational Technology Systems*, Vol. 16, No. 2, 1987–1988, pp. 151–169.

Holmgren, J. E., F. N. Dyer, R. E. Hilligoss, and F. H. Heller, "Effectiveness of Army Training Extension Course Lessons on Videodisk," *Journal of Educational Technology Systems*, Vol. 8, No. 3, 1979–1980, pp. 263–274.

Johnson, L. A., A. P. Jones, M. C. Butler, and D. Mann, *Assessing Interrater Agreement in Job Analysis Ratings*, Naval Health Research Center, Report No. 81-17, San Diego, California, 1981.

Ketner, W. D., "The Videodisc/Microprocessor for Training," *Training and Development Journal*, May 1981, pp. 151–153.

Kulik, Chen-Lin C., J. A. Kulik, and B. J. Schwalb, "The Effectiveness of Computer-Based Adult Education: A Meta-analysis," *Journal of Educational Computing Research*, 1986.

Kulik, J. A., Chen-Lin C. Kulik, and R. L. Bangert-Drowns, "Effectiveness of Computer-Based Education in Elementary Schools," *Computers in Human Behavior*, Vol. 1, 1985, pp. 59–74.

Kulik, James A., Chen-Lin C. Kulik, and Peter A. Cohen, "Effectiveness of Computer-based College Teaching: A Meta-analysis of Findings," *Review of Educational Research*, Vol. 50, No. 4, Winter 1980, pp. 525–544.

Maier, M. H., and C. M. Hiatt, *On the Content and Measurement Validity of Hands-On Job Performance Tests*, Center for Naval Analyses, CRM 85-79, Alexandria, Virginia, August 1985.

Manning, D. T., "Student Acceptance of Videodisk-Based Programs for Paramedical Training," *T.H.E. Journal*, Vol. 11, No. 3, 1983, pp. 105–108.

Niemic, Richard, and H. J. Walberg, "Comparative Effects of Computer-Assisted Instruction: A Synthesis of Reviews," *Journal of Educational Computing Research*, Vol. 3, No. 1, 1987, pp. 19–37.

Nugent, W. A., G. J. Laabs, and R. C. Panell, *Performance Test Objectivity: A Comparison of Rater Accuracy and Reliability Using Three Observation Forms*, Navy Personnel Research and

Development Center, NPRDC TR 82-30, San Diego, California, 1982.

Office of Technology Assessment, U.S. Congress, *Strategies for Medical Technology Assessment*, U.S. Government Printing Office, Washington, D.C., 1982.

Orlansky, J., "Effectiveness of CAI: A Different Finding," *Electronic Learning*, Vol. 3, No. 1, September 1983, pp. 58–60.

Orlansky, J., and J. String, "Computer-based Instruction for Military Training," *Defense Management Journal*, 1981, pp. 46–54.

Orlansky, J., and J. String, *Cost-Effectiveness of Computer-Based Instruction in Military Training*, Institute for Defense Analysis, IDA Paper P-1375, April 1979.

Polich, J. Michael, James N. Dertouzos, and S. James Press, *The Enlistment Bonus Experiment*, The RAND Corporation, R-3353-FMP, 1986.

Press, S. James, "The MISER Criterion for Imbalance in the Analysis of Covariance," *Journal of Statistical Planning and Inference*, Vol. 17, 1987, pp. 357–388.

Salomon, G., and H. Gardner, "The Computer as Educator: Lessons from Television Research," *Educational Researcher*, Vol. 15, No. 1, January 1986, pp. 13–19.

Scheffé, Henry, *The Analysis of Variance*, John Wiley and Sons, New York, 1959.

Shavelson, R. J., C. Stasz, S. Schlossman, N. Webb, J. Hotta, and S. Goldstein, *Evaluating Student Outcomes from Telecourse Instruction: A Feasibility Study*, The RAND Corporation, R-3422-CPB, May 1986.

Shlechter, Theodore M., *An Examination of the Research Evidence for Computer-Based Instruction in Military Training*, U.S. Army Research Institute for the Behavioral and Social Sciences, ARI Field Unit, Fort Knox, Kentucky, August 1986.

Vernon, C. D., *Evaluation of Interactive Video Disc System for Training the Operation of the DCT-9000 in the MOS 72G Course*, U.S. Army Communicative Technology Office, Report TR-84-6, October 1984.

Vinsonhaler, John F., and R. K. Bass, "A Summary of Ten Major Studies on CAI Drill and Practice," *Educational Technology*, July 1972, pp. 29–37.

Wilkinson, G. L., *An Evaluation of Equipment—Independent Maintenance Training by Means of a Microprocessor-Controlled Videodisc Delivery System*, Battelle Memorial Institute, Report TDI-TR-83-1, March 1983.

Wilkinson, G. L., *Evaluation of the Effectiveness and Potential Application of the Interactive Videodisc Instructional Delivery System within the Training of SIGINT/EW Systems Repairers*, Battelle Memorial Institute, Report TRS-85-1, March 1985.

Winer, B. J., *Statistical Principles in Experimental Design*, 2nd edition, McGraw-Hill, New York, 1971.

Winkler, J. D., "Army Applications of Interactive Videodisc Technology," The RAND Corporation, forthcoming.

Yeaton, W. H., and L. Sechrest, "Assessing Factors Influencing Acceptance of No-Difference Research," *Evaluation Review*, Vol. 11, No. 1, February 1987, pp. 131–142.

Young, J. I., and D. T. Tosti, *Equipment-Independent Training Program*, U.S. Army Training and Doctrine Command, Training Developments Institute, Report TDI-TR-3-81, December 1981.